

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 50.2179 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221b-10

Perfect score: 1340

Sequence: 1 GLKGRGDSGSPATWTTRGF.....KAGELEKTIISRCQVCMKKRH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1670	1 CGH03B	collagen alpha 3(I)
2	1218.5	90.9	246	2 I48302	collagen alpha 3(I)
3	1210.5	90.3	471	2 A39024	collagen alpha 3(I)
4	960	71.6	1669	1 CGH04B	collagen alpha 1(I)
5	952.5	71.1	253	2 I48304	collagen alpha 5(I)
6	951	71.0	1669	1 CGM34B	collagen alpha 1(I)
7	943.5	70.4	1691	1 S22917	collagen alpha 1(I)
8	943	70.4	258	2 B61228	collagen alpha 1(I)
9	922.5	68.8	754	2 A55267	collagen alpha 5(I)
10	872	65.1	220	2 B49736	collagen alpha 3(I)
11	852.5	63.6	1744	2 S40991	collagen alpha 1(I)
12	834	62.2	161	2 S49488	collagen alpha 3(I)
13	832	62.1	1782	2 A54507	collagen alpha 3(I)
14	796.5	59.4	1783	2 S16366	collagen alpha 1(I)
15	783.5	58.5	1758	2 T29350	collagen alpha 5(I)
16	783.5	58.5	1759	2 T29351	collagen alpha 2(I)
17	782	58.4	261	2 A34476	collagen alpha 2(I)
18	767.5	57.3	1747	2 A54121	collagen alpha-4 C
19	760.5	56.8	1707	2 A33526	collagen alpha 2(I)
20	755.5	56.4	1712	1 CGH02B	collagen alpha 2(I)
21	739	55.1	1691	1 CGH06B	collagen alpha 6(I)
22	734.5	54.8	775	2 A61228	collagen alpha 2(I)
23	709	52.9	453	2 S18804	collagen alpha 4(I)
24	705	52.6	312	2 I48303	collagen alpha 4(I)
25	702	52.4	623	2 A45137	collagen alpha 4(I)
26	696	51.9	1690	1 CGH01B	collagen alpha 4(I)
27	694.5	51.8	1775	2 A31893	collagen alpha 1(I)
28	646.5	48.2	1761	2 T13990	collagen type IV a
29	334	24.9	81	2 A49736	collagen alpha 3(I)

## ALIGNMENTS

## RESULT 1

## CGH03B

collagen alpha 3(IV) chain precursor, long splice form - human

N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form

C/Species: Homo sapiens (man)

C/Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999

C/Accession: A54763; A43928; A4043; A45971; A39786

R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Readers, S.T.

J. Biol. Chem. 269, 23013-23017, 1994

A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression

A/Reference number: A54763; MUID:94364994; PMID:8083201

A/Accession: A54763

A/Molecule type: mRNA

A/Residues: 1-1670 <MAR>

A/Cross-references: GB:X80031; NID:G577563; PID:G577564

A/Experimental source: kidney

R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the

A/Reference number: A43928; MUID:92147878; PMID:1737849

A/Accession: A43928

A/Molecule type: mRNA

A/Residues: 1331-1524, 1, 1526-1670 <TUR>

A/Cross-references: GB:M81379

A/Experimental source: kidney

R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 19780-19784, 1992

A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture

A/Reference number: A44043; MUID:93015826; PMID:1400291

A/Accession: A44043

A/Molecule type: DNA; mRNA

A/Residues: 1386-1670 <QUI>

A/Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896

A/Note: sequence extracted from NCEI backbone (NCBIP:115597)

R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 17358, 1994

A/Reference number: A44738; MUID:94274734; PMID:8006044

A/Contents: annotation; erratum; correction to intronic sequence in A44043

R/Bernal, D.; Quinones, S.; Saus, J.

J. Biol. Chem. 268, 12090-12094, 1993

A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

A/Reference number: A45971; MUID:93280184; PMID:8505332

A/Accession: A45971

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1427-1444 <BER>

A/Note: sequence extracted from NCEI backbone (NCBIP:133363); sequence incorrectly ide

R/Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Readers, S.T.

Am. J. Hum. Genet. 49, 545-554, 1991

A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain o

A/Reference number: A39786; MUID:91353570; PMID:1882840



A;Accession: A39786  
A;Molecule type: mRNA  
A;Residues: 1453-1593, 'A', 1595-1670 <MOR>  
A;Cross-references: GB:S55790; NID:Q234418; PIDN:AA819637.1; PID:G234419  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
A;Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3(IV)  
monomer trimers amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-  
C;Function:  
A;Description: minor structural component of extracellular basement membrane in kidney  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
F;1-28/Domain: signal sequence #status predicted <SIG>  
F;29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>  
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>  
F;43-1438/Region: interrupted helical  
F;791-793/Region: cell attachment (R-G-D) motif  
F;996-998/Region: cell attachment (R-G-D) motif  
F;1154-1156/Region: cell attachment (R-G-D) motif  
F;1306-1308/Region: cell attachment (R-G-D) motif  
F;1345-1347/Region: cell attachment (R-G-D) motif  
F;1432-1434/Region: cell attachment (R-G-D) motif  
F;1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F;161-33,39,41,125,422,476,479,682,722,809,1387/D:disulfide bonds: interchain #status predi  
F;253/Binding site: carboxylate (Asn) (covalent) #status predicted  
F;1460-1548,1493-1551/disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F;1505-1511,1616-1622/disulfide bonds: #status predicted  
F;1570-1662,1604-1665/disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
Query Match 100.0%; Score 1340; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 4.8e-113;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60  
DB 1427 GLKGRGDSGSPATWTRGFTVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 1486  
QY 61 LGTLGSLQRFMTTTPFLFCNVNDVCFASRNDSYSLWSTPPLMPNMNAPITGRLAEPYIS 120  
DB 1487 LGTLGSLQRFMTTTPFLFCNVNDVCFASRNDSYSLWSTPPLMPNMNAPITGRLAEPYIS 1546  
QY 121 RCTVCGPATAIAVHSQTTDIPPCPHGWTSLWKGFSLFMTSAGSEGTGQALASPGSCLE 180  
DB 1547 RCTVCGPATAIAVHSQTTDIPPCPHGWTSLWKGFSLFMTSAGSEGTGQALASPGSCLE 1606  
QY 181 EFRASPLFCHGRCNTYNSYSFSLWASLNPERFRKPIPTSTVKGLEKIISRCQVC 240  
DB 1607 EFRASPLFCHGRCNTYNSYSFSLWASLNPERFRKPIPTSTVKGLEKIISRCQVC 1666  
QY 241 MKRH 244  
DB 1667 MKRH 1670  
RESULT 2  
collagen alpha 3(IV) chain - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
C;Accession: 148302; S47278  
R;Miner, J.H.; Sanes, J.R.  
J. Cell Biol. 127, 879-894, 1994  
A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A;Reference number: A54979; MUID:95050957; PMID:7962065  
A;Accession: 148302  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-246 <RES>  
A;Cross-references: EMBL:Z35166; NID:G535197; PID:G535198  
C;Superfamily: collagen alpha 1(IV) chain  
Query Match 90.9%; Score 1218.5; DB 2; Length 246;  
Best Local Similarity 90.3%; Pred. No. 6.1e-103;  
Matches 221; Conservative 12; Mismatches 11; Indels 1; Gaps 1;  
QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQ 59  
DB 2 GLKGNPCDRTGPTATGTRGFTVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQ 61  
QY 60 DLGTLGSLQRFMTTTPFLFCNVNDVCFASRNDSYSLWSTPPLMPNMNAPITGRLAEPYI 119  
DB 62 DLGTLGSLQRFMTTTPFLFCNVNDVCFASRNDSYSLWSTPPLMPNMNAPITGRLAEPYI 121  
QY 120 SRTVCGPATAIAVHSQTTDIPPCPHGWTSLWKGFSLFMTSAGSEGTGQALASPGSCL 179  
DB 122 SRTVCGPATAIAVHSQTTAIPCCQDWSLWKGFSLFMTSAGSEGTGQALASPGSCL 181  
QY 180 EFRASPLFCHGRCNTYNSYSFSLWASLNPERFRKPIPTSTVKGLEKIISRCQVC 239  
DB 182 EFRASPLFCHGRCNTYNSYSFSLWASLNPERFRKPIPTSTVKGLEKIISRCQVC 241  
QY 240 MKRH 244  
DB 242 MKRH 246  
RESULT 3  
A39024  
collagen alpha 3(IV) chain - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815  
R;Morrisson, K.E.; Germino, G.G.; Reiders, S.T.  
J. Biol. Chem. 266, 34-39, 1991  
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
A;Reference number: A39024; MUID:91093146; PMID:1985905  
A;Accession: A39024  
A;Molecule type: mRNA  
A;Residues: 1-471 <MOR>  
A;Cross-references: EMBL:M63139; NID:G162886; PIDN:AAA62708.1; PID:G162887  
R;Burkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
A;Reference number: S18432; MUID:87222419; PMID:2438283  
A;Accession: S20672  
A;Molecule type: protein  
A;Residues: 237-238, 'X', 230-244 <BUT>  
R;Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
J. Biol. Chem. 263, 13374-13380, 1988  
A;Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
A;Reference number: S17802; MUID:88330844; PMID:3417561  
A;Accession: S17802  
A;Molecule type: protein  
A;Residues: 237-238, 'X', 230-252, 'Y', 254 <SAU>  
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5466-5469, 1990  
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
A;Reference number: A35167; MUID:90202779; PMID:2318822  
A;Accession: A35167  
A;Molecule type: protein  
A;Residues: 236-258 <GUN>  
R;Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Mc  
J. Biol. Chem. 266, 15318-15324, 1991  
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol  
A;Reference number: A39419; MUID:91332055; PMID:1869555  
A;Accession: C39419

A:Molecule type: protein  
A:Residues: 236-255 <G02>  
A:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F:1-238/Domain: collagenous (fragment) #status predicted <COL>  
F:239-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC11>  
F:239-353/Domain: repeat NCI #status predicted <NC11>  
F:354-471/Domain: repeat NCI #status predicted <NC12>  
F:232,238/Modified site: hydroxyproline (Pro) #status experimental  
F:306-312,417-423/disulfide bonds: #status predicted

Query Match 90.3%; Score 1210.5; DB 2; Length 471;  
Best Local Similarity 90.6%; Pred. No. 6.5e-102;  
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRKSGSPATWT-RGFVTRHSQTTA-PSCEGVPLYSFGFSLFVQGNORAGQ 59  
DB 227 GLKGRKSGSPATWT-RGFVTRHSQTTA-PSCEGVPLYSFGFSLFVQGNORAGQ 286  
QY 60 DLGTLGSLQRTTTPFLFCNVNDYCNFASNDYSYWLSTPALMPNMAPITGRALEPYI 119  
DB 287 DLGTLGSLQRTTTPFLFCNVNDYCNFASNDYSYWLSTPALMPNMAPITGRALEPYI 346  
QY 120 SRTCTCEGPAIAIVHSQTTDTPCPHGWSISWKGFSPMTFSAGSEGTGOALSPGSL 179  
DB 347 SRTCTCEGPAIAIVHSQTTDTPCPHGWSISWKGFSPMTFSAGSEGTGOALSPGSL 406  
QY 180 EEFRAFPFTECHGRGFCNYNSYSFPLASLNPFRMFKPIPTVKAGELEKIISRCQVC 239  
DB 407 EEFRAFPFTECHGRGFCNYNSYSFPLASLNPFRMFKPIPTVKAGELEKIISRCQVC 466  
QY 240 MKKR 243  
DB 467 MKKR 470

RESULT 4  
CGHU4B  
collagen alpha 1(IV) chain precursor - human  
N:Alternate names: procollagen alpha 1(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58  
R:Scioinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV coll  
A:Reference number: S16876; MUID:89340433; PMID:2701944  
A:Accession: S16876  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1669 <SO11>  
A:Cross-references: EMBL:J04217; GB:J05039; NID:G180759; PIDN:AAAS3097.1; PID:G553233  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R:Scioinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A:Reference number: A92690; MUID:89034231; PMID:3182844  
A:Accession: A32117  
A:Molecule type: DNA  
A:Residues: 1-28 <SO12>  
A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G553233  
R:Poerschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c  
A:Reference number: S02738; MUID:89030632; PMID:2846280  
A:Accession: S02738  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-6, 'L', 8-28 <POE>  
A:Cross-references: EMBL:X12784; NID:G30072  
R:Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutmann, R.;  
Eur. J. Biochem. 168, 529-536, 1987  
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem

A:Reference number: S00048; MUID:88029471; PMID:3311751  
A:Accession: S00048  
A:Molecule type: mRNA  
A:Residues: 1-318, 'A', 320-944 <BRA1>  
A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A:Accession: S25826  
A:Molecule type: protein  
A:Residues: 271-318, 'A', 320-554 <BRA2>  
R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7.  
A:Reference number: A23115; MUID:86004708; PMID:4043082  
A:Accession: A23115  
A:Molecule type: protein  
A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A:Experimental source: placenta  
A:Note: the amino end of the mature form is blocked  
R:Scioinen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1987  
A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (.  
A:Reference number: S00207; MUID:88083584; PMID:3691802  
A:Accession: S00207  
A:Molecule type: mRNA  
A:Residues: 244-530 <SO13>  
A:Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
R:Edle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collage  
A:Reference number: S39614; MUID:94038963; PMID:8223488  
A:Accession: S39614  
A:Molecule type: protein  
A:Residues: 371-554 <EBL>  
R:Babel, W.; Glanville, R.W.  
Eur. J. Biochem. 143, 545-556, 1984  
A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid  
A:Reference number: A02863; MUID:85003629; PMID:6434307  
A:Accession: A02863  
A:Molecule type: protein  
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99  
A:Experimental source: placenta  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A:Title: Pepsin fragments of human placental basement-membrane collagens showing inter  
A:Reference number: S16908; MUID:82005835; PMID:6792033  
A:Accession: A58517  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-  
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane  
A:Reference number: S16910; MUID:84033346; PMID:6416291  
A:Accession: S16910  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'Y', 946, 'X', 94  
A:Experimental source: placenta  
R:Philaianemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;  
J. Biol. Chem. 260, 7681-7687, 1985  
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen  
A:Reference number: S01466; MUID:85207819; PMID:2581969  
A:Accession: S01466  
A:Molecule type: mRNA  
A:Residues: 1256-1669 <PTH>  
A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424  
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chain  
A:Reference number: S16879; MUID:85216555; PMID:2582422  
A:Accession: S16879  
A:Molecule type: mRNA  
A:Residues: 1259-1669 <BRI>  
A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAAS2042.1; PID:G180818  
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss  
Eur. J. Biochem. 147, 217-224, 1985



A;Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-711  
A;Cross-references: EMBL:X06777  
R;Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.  
J. Biol. Chem. 263, 8706-8709, 1988  
A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
A;Reference number: A28066; MUID:88243724; PMID:3379041  
A;Accession: A28066  
A;Molecule type: mRNA  
A;Residues: 1-129 <K1>  
A;Cross-references: EMBL:J03758; NID:G192669; PIDN:AAA37439.1; PID:G192670  
R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985  
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1(IV) collagen chain  
A;Reference number: A02864; MUID:85127033; PMID:2578961  
A;Accession: A02864  
A;Molecule type: mRNA  
A;Residues: 1276-1669 <OB>  
A;Cross-references: EMBL:X02201; NID:G50233; PIDN:CRA26132.1; PID:G1333876  
R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.  
Gene 43, 301-304, 1986  
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeoxynucleotide  
A;Reference number: A25636; MUID:86301886; PMID:3755692  
A;Accession: A25636  
A;Molecule type: mRNA  
A;Residues: 1149-1396, 'S', 1398-1424 <NA>  
A;Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287  
A;Note: the authors translated the codon CAG for residue 1374 as Arg  
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlajä J. Biol. Chem. 262, 8496-8499, 1987  
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV) collagen and human alpha-1(IV) collagen  
A;Reference number: A94680; MUID:87250460; PMID:3597383  
A;Accession: A29301  
A;Molecule type: mRNA  
A;Residues: 1441-1669 <KUR>  
A;Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115  
R;Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y.  
J. Biol. Chem. 263, 12310-12314, 1988  
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequence  
A;Reference number: S19079; MUID:88315019; PMID:2842328  
A;Accession: S19079  
A;Molecule type: DNA  
A;Residues: 1-28 <K12>  
A;Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503  
R;Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
J. Biol. Chem. 263, 19274-19277, 1988  
A;Title: Head-to-head arrangement of murine type IV collagen genes.  
A;Reference number: A92702; MUID:89066738; PMID:3198626  
A;Accession: A32003  
A;Molecule type: DNA  
A;Residues: 1-28 <KAY>  
A;Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449  
R;Burbelo, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promoter  
A;Reference number: A94220; MUID:89071759; PMID:3200851  
A;Accession: A31766  
A;Molecule type: DNA  
A;Residues: 1-28 <BUR>  
A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668  
R;Sakurai, Y.; Sullivan, M.; Yamada, Y.  
J. Biol. Chem. 261, 6654-6657, 1986  
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
A;Reference number: S19094; MUID:86196099; PMID:3009468  
A;Accession: S19094  
A;Molecule type: DNA  
A;Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>  
A;Cross-references: EMBL:M13027  
R;Schuppan, D.; Timpl, R.; Gnanville, R.W.  
FEBS Lett. 115, 297-300, 1980  
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (type IV) collagen  
A;Reference number: S16909; MUID:80246483; PMID:6772473  
A;Accession: S16909  
A;Molecule type: protein

A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-1237, 'X', 1239-1240, 'G', 1242-1243, 'X', 1245-1246, 'G', 1248-1249, 'X', 1251-1252, 'G', 1254-1255, 'X', 1257-1258, 'G', 1260-1261, 'X', 1263-1264, 'G', 1266-1267, 'X', 1269-1270, 'G', 1272-1273, 'X', 1275-1276, 'G', 1278-1279, 'X', 1281-1282, 'G', 1284-1285, 'X', 1287-1288, 'G', 1290-1291, 'X', 1293-1294, 'G', 1296-1297, 'X', 1299-1300, 'G', 1302-1303, 'X', 1305-1306, 'G', 1308-1309, 'X', 1311-1312, 'G', 1314-1315, 'X', 1317-1318, 'G', 1320-1321, 'X', 1323-1324, 'G', 1326-1327, 'X', 1329-1330, 'G', 1332-1333, 'X', 1335-1336, 'G', 1338-1339, 'X', 1341-1342, 'G', 1344-1345, 'X', 1347-1348, 'G', 1350-1351, 'X', 1353-1354, 'G', 1356-1357, 'X', 1359-1360, 'G', 1362-1363, 'X', 1365-1366, 'G', 1368-1369, 'X', 1371-1372, 'G', 1374-1375, 'X', 1377-1378, 'G', 1380-1381, 'X', 1383-1384, 'G', 1386-1387, 'X', 1389-1390, 'G', 1392-1393, 'X', 1395-1396, 'G', 1398-1399, 'X', 1401-1402, 'G', 1404-1405, 'X', 1407-1408, 'G', 1410-1411, 'X', 1413-1414, 'G', 1416-1417, 'X', 1419-1420, 'G', 1422-1423, 'X', 1425-1426, 'G', 1428-1429, 'X', 1431-1432, 'G', 1434-1435, 'X', 1437-1438, 'G', 1440-1441, 'X', 1443-1444, 'G', 1446-1447, 'X', 1449-1450, 'G', 1452-1453, 'X', 1455-1456, 'G', 1458-1459, 'X', 1461-1462, 'G', 1464-1465, 'X', 1467-1468, 'G', 1470-1471, 'X', 1473-1474, 'G', 1476-1477, 'X', 1479-1480, 'G', 1482-1483, 'X', 1485-1486, 'G', 1488-1489, 'X', 1491-1492, 'G', 1494-1495, 'X', 1497-1498, 'G', 1500-1501, 'X', 1503-1504, 'G', 1506-1507, 'X', 1509-1510, 'G', 1512-1513, 'X', 1515-1516, 'G', 1518-1519, 'X', 1521-1522, 'G', 1524-1525, 'X', 1527-1528, 'G', 1530-1531, 'X', 1533-1534, 'G', 1536-1537, 'X', 1539-1540, 'G', 1542-1543, 'X', 1545-1546, 'G', 1548-1549, 'X', 1551-1552, 'G', 1554-1555, 'X', 1557-1558, 'G', 1560-1561, 'X', 1563-1564, 'G', 1566-1567, 'X', 1569-1570, 'G', 1572-1573, 'X', 1575-1576, 'G', 1578-1579, 'X', 1581-1582, 'G', 1584-1585, 'X', 1587-1588, 'G', 1590-1591, 'X', 1593-1594, 'G', 1596-1597, 'X', 1599-1600, 'G', 1602-1603, 'X', 1605-1606, 'G', 1608-1609, 'X', 1611-1612, 'G', 1614-1615, 'X', 1617-1618, 'G', 1620-1621, 'X', 1623-1624, 'G', 1626-1627, 'X', 1629-1630, 'G', 1632-1633, 'X', 1635-1636, 'G', 1638-1639, 'X', 1641-1642, 'G', 1644-1645, 'X', 1647-1648, 'G', 1650-1651, 'X', 1653-1654, 'G', 1656-1657, 'X', 1659-1660, 'G', 1662-1663, 'X', 1665-1666, 'G', 1668-1669, 'X', 1671-1672, 'G', 1674-1675, 'X', 1677-1678, 'G', 1680-1681, 'X', 1683-1684, 'G', 1686-1687, 'X', 1689-1690, 'G', 1692-1693, 'X', 1695-1696, 'G', 1698-1699, 'X', 1701-1702, 'G', 1704-1705, 'X', 1707-1708, 'G', 1710-1711, 'X', 1713-1714, 'G', 1716-1717, 'X', 1719-1720, 'G', 1722-1723, 'X', 1725-1726, 'G', 1728-1729, 'X', 1731-1732, 'G', 1734-1735, 'X', 1737-1738, 'G', 1740-1741, 'X', 1743-1744, 'G', 1746-1747, 'X', 1749-1750, 'G', 1752-1753, 'X', 1755-1756, 'G', 1758-1759, 'X', 1761-1762, 'G', 1764-1765, 'X', 1767-1768, 'G', 1770-1771, 'X', 1773-1774, 'G', 1776-1777, 'X', 1779-1780, 'G', 1782-1783, 'X', 1785-1786, 'G', 1788-1789, 'X', 1791-1792, 'G', 1794-1795, 'X', 1797-1798, 'G', 1800-1801, 'X', 1803-1804, 'G', 1806-1807, 'X', 1809-1810, 'G', 1812-1813, 'X', 1815-1816, 'G', 1818-1819, 'X', 1821-1822, 'G', 1824-1825, 'X', 1827-1828, 'G', 1830-1831, 'X', 1833-1834, 'G', 1836-1837, 'X', 1839-1840, 'G', 1842-1843, 'X', 1845-1846, 'G', 1848-1849, 'X', 1851-1852, 'G', 1854-1855, 'X', 1857-1858, 'G', 1860-1861, 'X', 1863-1864, 'G', 1866-1867, 'X', 1869-1870, 'G', 1872-1873, 'X', 1875-1876, 'G', 1878-1879, 'X', 1881-1882, 'G', 1884-1885, 'X', 1887-1888, 'G', 1890-1891, 'X', 1893-1894, 'G', 1896-1897, 'X', 1899-1900, 'G', 1902-1903, 'X', 1905-1906, 'G', 1908-1909, 'X', 1911-1912, 'G', 1914-1915, 'X', 1917-1918, 'G', 1920-1921, 'X', 1923-1924, 'G', 1926-1927, 'X', 1929-1930, 'G', 1932-1933, 'X', 1935-1936, 'G', 1938-1939, 'X', 1941-1942, 'G', 1944-1945, 'X', 1947-1948, 'G', 1950-1951, 'X', 1953-1954, 'G', 1956-1957, 'X', 1959-1960, 'G', 1962-1963, 'X', 1965-1966, 'G', 1968-1969, 'X', 1971-1972, 'G', 1974-1975, 'X', 1977-1978, 'G', 1980-1981, 'X', 1983-1984, 'G', 1986-1987, 'X', 1989-1990, 'G', 1992-1993, 'X', 1995-1996, 'G', 1998-1999, 'X', 2001-2002, 'G', 2004-2005, 'X', 2007-2008, 'G', 2010-2011, 'X', 2013-2014, 'G', 2016-2017, 'X', 2019-2020, 'G', 2022-2023, 'X', 2025-2026, 'G', 2028-2029, 'X', 2031-2032, 'G', 2034-2035, 'X', 2037-2038, 'G', 2040-2041, 'X', 2043-2044, 'G', 2046-2047, 'X', 2049-2050, 'G', 2052-2053, 'X', 2055-2056, 'G', 2058-2059, 'X', 2061-2062, 'G', 2064-2065, 'X', 2067-2068, 'G', 2070-2071, 'X', 2073-2074, 'G', 2076-2077, 'X', 2079-2080, 'G', 2082-2083, 'X', 2085-2086, 'G', 2088-2089, 'X', 2091-2092, 'G', 2094-2095, 'X', 2097-2098, 'G', 2100-2101, 'X', 2103-2104, 'G', 2106-2107, 'X', 2109-2110, 'G', 2112-2113, 'X', 2115-2116, 'G', 2118-2119, 'X', 2121-2122, 'G', 2124-2125, 'X', 2127-2128, 'G', 2130-2131, 'X', 2133-2134, 'G', 2136-2137, 'X', 2139-2140, 'G', 2142-2143, 'X', 2145-2146, 'G', 2148-2149, 'X', 2151-2152, 'G', 2154-2155, 'X', 2157-2158, 'G', 2160-2161, 'X', 2163-2164, 'G', 2166-2167, 'X', 2169-2170, 'G', 2172-2173, 'X', 2175-2176, 'G', 2178-2179, 'X', 2181-2182, 'G', 2184-2185, 'X', 2187-2188, 'G', 2190-2191, 'X', 2193-2194, 'G', 2196-2197, 'X', 2199-2200, 'G', 2202-2203, 'X', 2205-2206, 'G', 2208-2209, 'X', 2211-2212, 'G', 2214-2215, 'X', 2217-2218, 'G', 2220-2221, 'X', 2223-2224, 'G', 2226-2227, 'X', 2229-2230, 'G', 2232-2233, 'X', 2235-2236, 'G', 2238-2239, 'X', 2241-2242, 'G', 2244-2245, 'X', 2247-2248, 'G', 2250-2251, 'X', 2253-2254, 'G', 2256-2257, 'X', 2259-2260, 'G', 2262-2263, 'X', 2265-2266, 'G', 2268-2269, 'X', 2271-2272, 'G', 2274-2275, 'X', 2277-2278, 'G', 2280-2281, 'X', 2283-2284, 'G', 2286-2287, 'X', 2289-2290, 'G', 2292-2293, 'X', 2295-2296, 'G', 2298-2299, 'X', 2301-2302, 'G', 2304-2305, 'X', 2307-2308, 'G', 2310-2311, 'X', 2313-2314, 'G', 2316-2317, 'X', 2319-2320, 'G', 2322-2323, 'X', 2325-2326, 'G', 2328-2329, 'X', 2331-2332, 'G', 2334-2335, 'X', 2337-2338, 'G', 2340-2341, 'X', 2343-2344, 'G', 2346-2347, 'X', 2349-2350, 'G', 2352-2353, 'X', 2355-2356, 'G', 2358-2359, 'X', 2361-2362, 'G', 2364-2365, 'X', 2367-2368, 'G', 2370-2371, 'X', 2373-2374, 'G', 2376-2377, 'X', 2379-2380, 'G', 2382-2383, 'X', 2385-2386, 'G', 2388-2389, 'X', 2391-2392, 'G', 2394-2395, 'X', 2397-2398, 'G', 2400-2401, 'X', 2403-2404, 'G', 2406-2407, 'X', 2409-2410, 'G', 2412-2413, 'X', 2415-2416, 'G', 2418-2419, 'X', 2421-2422, 'G', 2424-2425, 'X', 2427-2428, 'G', 2430-2431, 'X', 2433-2434, 'G', 2436-2437, 'X', 2439-2440, 'G', 2442-2443, 'X', 2445-2446, 'G', 2448-2449, 'X', 2451-2452, 'G', 2454-2455, 'X', 2457-2458, 'G', 2460-2461, 'X', 2463-2464, 'G', 2466-2467, 'X', 2469-2470, 'G', 2472-2473, 'X', 2475-2476, 'G', 2478-2479, 'X', 2481-2482, 'G', 2484-2485, 'X', 2487-2488, 'G', 2490-2491, 'X', 2493-2494, 'G', 2496-2497, 'X', 2499-2500, 'G', 2502-2503, 'X', 2505-2506, 'G', 2508-2509, 'X', 2511-2512, 'G', 2514-2515, 'X', 2517-2518, 'G', 2520-2521, 'X', 2523-2524, 'G', 2526-2527, 'X', 2529-2530, 'G', 2532-2533, 'X', 2535-2536, 'G', 2538-2539, 'X', 2541-2542, 'G', 2544-2545, 'X', 2547-2548, 'G', 2550-2551, 'X', 2553-2554, 'G', 2556-2557, 'X', 2559-2560, 'G', 2562-2563, 'X', 2565-2566, 'G', 2568-2569, 'X', 2571-2572, 'G', 2574-2575, 'X', 2577-2578, 'G', 2580-2581, 'X', 2583-2584, 'G', 2586-2587, 'X', 2589-2590, 'G', 2592-2593, 'X', 2595-2596, 'G', 2598-2599, 'X', 2601-2602, 'G', 2604-2605, 'X', 2607-2608, 'G', 2610-2611, 'X', 2613-2614, 'G', 2616-2617, 'X', 2619-2620, 'G', 2622-2623, 'X', 2625-2626, 'G', 2628-2629, 'X', 2631-2632, 'G', 2634-2635, 'X', 2637-2638, 'G', 2640-2641, 'X', 2643-2644, 'G', 2646-2647, 'X', 2649-2650, 'G', 2652-2653, 'X', 2655-2656, 'G', 2658-2659, 'X', 2661-2662, 'G', 2664-2665, 'X', 2667-2668, 'G', 2670-2671, 'X', 2673-2674, 'G', 2676-2677, 'X', 2679-2680, 'G', 2682-2683, 'X', 2685-2686, 'G', 2688-2689, 'X', 2691-2692, 'G', 2694-2695, 'X', 2697-2698, 'G', 2700-2701, 'X', 2703-2704, 'G', 2706-2707, 'X', 2709-2710, 'G', 2712-2713, 'X', 2715-2716, 'G', 2718-2719, 'X', 2721-2722, 'G', 2724-2725, 'X', 2727-2728, 'G', 2730-2731, 'X', 2733-2734, 'G', 2736-2737, 'X', 2739-2740, 'G', 2742-2743, 'X', 2745-2746, 'G', 2748-2749, 'X', 2751-2752, 'G', 2754-2755, 'X', 2757-2758, 'G', 2760-2761, 'X', 2763-2764, 'G', 2766-2767, 'X', 2769-2770, 'G', 2772-2773, 'X', 2775-2776, 'G', 2778-2779, 'X', 2781-2782, 'G', 2784-2785, 'X', 2787-2788, 'G', 2790-2791, 'X', 2793-2794, 'G', 2796-2797, 'X', 2799-2800, 'G', 2802-2803, 'X', 2805-2806, 'G', 2808-2809, 'X', 2811-2812, 'G', 2814-2815, 'X', 2817-2818, 'G', 2820-2821, 'X', 2823-2824, 'G', 2826-2827, 'X', 2829-2830, 'G', 2832-2833, 'X', 2835-2836, 'G', 2838-2839, 'X', 2841-2842, 'G', 2844-2845, 'X', 2847-2848, 'G', 2850-2851, 'X', 2853-2854, 'G', 2856-2857, 'X', 2859-2860, 'G', 2862-2863, 'X', 2865-2866, 'G', 2868-2869, 'X', 2871-2872, 'G', 2874-2875, 'X', 2877-2878, 'G', 2880-2881, 'X', 2883-2884, 'G', 2886-2887, 'X', 2889-2890, 'G', 2892-2893, 'X', 2895-2896, 'G', 2898-2899, 'X', 2901-2902, 'G', 2904-2905, 'X', 2907-2908, 'G', 2910-2911, 'X', 2913-2914, 'G', 2916-2917, 'X', 2919-2920, 'G', 2922-2923, 'X', 2925-2926, 'G', 2928-2929, 'X', 2931-2932, 'G', 2934-2935, 'X', 2937-2938, 'G', 2940-2941, 'X', 2943-2944, 'G', 2946-2947, 'X', 2949-2950, 'G', 2952-2953, 'X', 2955-2956, 'G', 2958-2959, 'X', 2961-2962, 'G', 2964-2965, 'X', 2967-2968, 'G', 2970-2971, 'X', 2973-2974, 'G', 2976-2977, 'X', 2979-2980, 'G', 2982-2983, 'X', 2985-2986, 'G', 2988-2989, 'X', 2991-2992, 'G', 2994-2995, 'X', 2997-2998, 'G', 3000-3001, 'X', 3003-3004, 'G', 3006-3007, 'X', 3009-3010, 'G', 3012-3013, 'X', 3015-3016, 'G', 3018-3019, 'X', 3021-3022, 'G', 3024-3025, 'X', 3027-3028, 'G', 3030-3031, 'X', 3033-3034, 'G', 3036-3037, 'X', 3039-3040, 'G', 3042-3043, 'X', 3045-3046, 'G', 3048-3049, 'X', 3051-3052, 'G', 3054-3055, 'X', 3057-3058, 'G', 3060-3061, 'X', 3063-3064, 'G', 3066-3067, 'X', 3069-3070, 'G', 3072-3073, 'X', 3075-3076, 'G', 3078-3079, 'X', 3081-3082, 'G', 3084-3085, 'X', 3087-3088, 'G', 3090-3091, 'X', 3093-3094, 'G', 3096-3097, 'X', 3099-3100, 'G', 3102-3103, 'X', 3105-3106, 'G', 3108-3109, 'X', 3111-3112, 'G', 3114-3115, 'X', 3117-3118, 'G', 3120-3121, 'X', 3123-3124, 'G', 3126-3127, 'X', 3129-3130, 'G', 3132-3133, 'X', 3135-3136, 'G', 3138-3139, 'X', 3141-3142, 'G', 3144-3145, 'X', 3147-3148, 'G', 3150-3151, 'X', 3153-3154, 'G', 3156-3157, 'X', 3159-3160, 'G', 3162-3163, 'X', 3165-3166, 'G', 3168-3169, 'X', 3171-3172, 'G', 3174-3175, 'X', 3177-3178, 'G', 3180-3181, 'X', 3183-3184, 'G', 3186-3187, 'X', 3189-3190, 'G', 3192-3193, 'X', 3195-3196, 'G', 3198-3199, 'X', 3201-3202, 'G', 3204-3205, 'X', 3207-3208, 'G', 3210-3211, 'X', 3213-3214, 'G', 3216-3217, 'X', 3219-3220, 'G', 3222-3223, 'X', 3225-3226, 'G', 3228-3229, 'X', 3231-3232, 'G', 3234-3235, 'X', 3237-3238, 'G', 3240-3241, 'X', 3243-3244, 'G', 3246-3247, 'X', 3249-3250, 'G', 3252-3253, 'X', 3255-3256, 'G', 3258-3259, 'X', 3261-3262, 'G', 3264-3265, 'X', 3267-3268, 'G', 3270-3271, 'X', 3273-3274, 'G', 3276-3277, 'X', 3279-3280, 'G', 3282-3283, 'X', 3285-3286, 'G', 3288-3289, 'X', 3291-3292, 'G', 3294-3295, 'X', 3297-3298, 'G', 3300-3301, 'X', 3303-3304, 'G', 3306-3307, 'X', 3309-3310, 'G', 3312-3313, 'X', 3315-3316, 'G', 3318-3319, 'X', 3321-3322, 'G', 3324-3325, 'X', 3327-3328, 'G', 3330-3331, 'X', 3333-3334, 'G', 3336-3337, 'X', 3339-3340, 'G', 3342-3343, 'X', 3345-3346, 'G', 3348-3349, 'X', 3351-3352, 'G', 3354-3355, 'X', 3357-3358, 'G', 3360-3361, 'X', 3363-3364, 'G', 3366-3367, 'X', 3369-3370, 'G', 3372-3373, 'X', 3375-3376, 'G', 3378-3379, 'X', 3381-3382, 'G', 3384-3385, 'X', 3387-3388, 'G', 3390-3391, 'X', 3393-3394, 'G', 3396-3397, 'X', 3399-3400, 'G', 3402-3403, 'X', 3405-3406, 'G', 3408-3409, '

N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text\_change 21-Jul-2000  
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A39  
J: Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi  
n Alport syndrome patient.  
A:Reference number: S22917; MUID:92316923; PMID:1352287  
A:Accession: S22917  
A:Molecule type: mRNA  
A:Residues: 1-967 <ZHO>  
A:CROSS-references: GB:M90464; NID:q180826; PIDN:AAA52046.1; PID:g553234  
J: Zhou, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A:Title: Structure of the human type IV collagen COL4A5 gene.  
A:Reference number: A54365; MUID:94165049; PMID:8120014  
A:Accession: A54365  
A:Molecule type: DNA  
A:Residues: 1-922 <ZH2>  
A:CROSS-references: GB:U04470; NID:q463378; GB:U04520; NID:q463428; PIDN:AAC27816.1; PID  
R: Zhou, J.; Mochizuki, T.; Sneets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggvason  
Science 261, 1157-1169, 1993  
A:Title: Deletion of the paired alphas(IV) and alpha5(IV) collagen genes in inherited sc  
A:Reference number: A57079; MUID:93361972; PMID:8356449  
A:Accession: A57079  
A:Molecule type: DNA  
A:Residues: 1-27 <ZH4>  
A:CROSS-references: GB:Z37153; NID:q587203; PIDN:CAA85512.1; PID:g587204  
R: Pihlajaniemi, T.; Pihlajaniemi, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin  
A:Reference number: A37122; MUID:90337990; PMID:2380186  
A:Accession: A37122  
A:Molecule type: mRNA  
A:Residues: 84-439, 'GS', 442-624, 'LALO', 629-665, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>  
A:CROSS-references: GB:U05558; EMBL:M58526; NID:q134209  
A:Note: submitted to the EMBL Data Library, February 1991  
A:Note: the authors translated the codon GCC for residue 115 as Val  
R: Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma  
Hum. Mol. Genet. 1, 127-129, 1992  
A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in  
A:Reference number: I54317; MUID:93244772; PMID:1363780  
A:Accession: I54317  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 313-324, 'E', 326-330 <REN>  
A:CROSS-references: GB:S59334; NID:q299946; PIDN:AAD13909.1; PID:g4261609  
R: Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeythyaee, M.; Shows, T.B.; Tryggvason, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidne  
A:Reference number: A48450; MUID:90160375; PMID:1689491  
A:Accession: A48450  
A:Molecule type: mRNA  
A:Residues: 914-1264, 1271-1691 <HOS>  
A:CROSS-references: EMBL:M31115; NID:q180824; PIDN:AAA52045.1; PID:g180825  
R: Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
Genomics 9, 1-9, 1991  
A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene tha  
A:Reference number: A37969; MUID:91169491; PMID:2004755  
A:Accession: S18850  
A:Molecule type: DNA  
A:Residues: 924-1264, 1271-1691 <ZH3>  
A:CROSS-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB  
8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:q177922; PIDN:AAA51558.1; PID  
R: Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Casselman, J.J.; M  
Kidney Int. 44, 1316-1321, 1993  
A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex  
A:Reference number: I56971; MUID:94133540; PMID:8301933  
A:Accession: I56971  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1258-1276 <GUO1>

A:CROSS-references: GB:S63168; NID:g545095; PIDN:AAC60612.1; PID:g545096  
A:Note: kidney splice form  
A:Accession: I76598  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1284-1291, 'TFGLYLACLV' <GUO2>  
A:CROSS-references: GB:S69169; NID:g545097; PIDN:AAC60613.1; PID:g545098  
A:Note: frameshift mutation in patient with Alport syndrome  
R: Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; S  
Am. J. Hum. Genet. 46, 1024-1033, 1990  
A:Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the reg  
A:Reference number: A35335; MUID:90252791; PMID:2339699  
A:Accession: A35335  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1448-1477 <MYE>  
R: Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yc  
Kidney Int. 46, 1307-1314, 1994  
A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primor  
A:Reference number: I56975; MUID:95156893; PMID:7853788  
A:Accession: I56975  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1595-1602 <NAK>  
A:CROSS-references: GB:S75903; NID:q913882; PIDN:AAB33374.1; PID:g913883  
A:Note: premature termination mutation from a patient with Alport syndrome; one other n  
R: Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;  
Genomics 17, 485-489, 1993  
A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Al  
A:Reference number: I54188; MUID:94010948; PMID:8406498  
A:Accession: I54188  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1604-1607, 'VHDAYKC' <LEM>  
A:CROSS-references: GB:S65767; NID:q425563; PIDN:AAD13967.1; PID:g4261667  
A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations e  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: COL4A5; ATS  
A:CROSS-references: GDB:120596; OMIM:303630  
A:Map position: Xq22-Xq22  
A:Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 225  
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1;  
A:Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands wit  
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-l  
C:Function:  
A:Description: minor structural component of extracellular basement membrane  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; gly  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAI  
F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status  
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>  
F:27-1462/Region: interrupted helical  
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F:29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted  
F:125/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:1482-1573,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted  
F:1527-1533,1638-1644/Disulfide bonds: #status predicted  
F:1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted  
Query Match 70.4%; Score 943.5; DB 1; Length 1691;  
Best Local Similarity 67.9%; Pred. No. 4e-77;  
Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;  
Qy 1 GLKKGDCSGPATWT-TRGFVETRHSQTATPSCPEGVPLYSYSGFSLFVGNORAHQ 59  
Db 1448 GPDGLQGGPPPTSSVAHGLFTRHSQTADPCQCFGLQVYEGFSLLYVGNKRAHQ 1507

QY 60 DLGTLGSLQRTFTMPFLFCNVNDVNCNPNFASNDYSLWSTPALMPNNAPITGRALEPYI 119  
DB 1508 DLGTAGSCLRRFSTMPFECNINNVNPNFASNDYSLWSTPEPMPMSQPLKQSGIOFFI 1567  
QY 120 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCL 179  
DB 1568 SRCVACEAPAVVIAVHSQTIQIHPCPQGWDSLWATGYSEFMHTSAGSEGTGQALASPGSCL 1627  
QY 180 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMPKPIPTVTKAGELEKIISRCQVC 239  
DB 1628 EFRSAPFIECHGRGTCNYNSYSFWLATVDVSMFSKQSETLKAGDLRTRISRCQVC 1687  
QY 240 MKK 242  
DB 1688 MKR 1690

RESULT 8  
B61228  
collagen alpha 1(IV) chain - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 17-Mar-1999  
C:Accession: B61228  
R.Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.  
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991  
A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium  
A:Reference number: A61228; MUID:92010685; PMID:1717398  
A:Accession: B61228  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-258 <YAM>  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 70.4%; Score 943; DB 2; Length 258;  
Best Local Similarity 67.8%; Pred. No. 5.8e-78;  
Matches 164; Conservative 34; Mismatches 42; Indels 2; Gaps 1;  
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQD 60  
DB 18 GLPGSGMPGPTFS--VDHGFVTRHSQTTDHPQCPGKILYHGYSLLVQGNORAHQD 75  
QY 61 LGTLGSLQRTFTMPFLFCNVNDVNCNPNFASNDYSLWSTPALMPNNAPITGRALEPYIS 120  
DB 76 LGTAGSCLRRFSTMPFLFCNINNVNPNFASNDYSLWSTPEPMPMSQPLKQSGIOFFI 135  
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCL 180  
DB 136 RCVACEAPAVVIAVHSQTIQIHPCPQGWDSLWATGYSEFMHTSAGSEGTGQALASPGSCL 195  
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMPKPIPTVTKAGELEKIISRCQVC 240  
DB 196 EFRSAPFIECHGRGTCNYNSYSFWLATVDVSMFSKQSETLKAGDLRTRISRCQVC 255  
QY 241 KK 242  
DB 256 RR 257

RESULT 9  
A55267  
collagen alpha 5(IV) chain - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: A55267  
R.Zheng, K.; Thorne, P.S.; Marram, P.; Bauman, R.; McInnes, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994  
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked type IV.  
A:Reference number: A55267; MUID:94224868; PMID:8171024  
A:Accession: A55267  
A:Status: preliminary  
A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>  
A:Cross-references: GB:U07888; NID:9469547; PIDN:AA860258.1; PID:9469548  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 68.8%; Score 922.5; DB 2; Length 754;  
Best Local Similarity 68.4%; Pred. No. 1.3e-75;  
Matches 162; Conservative 32; Mismatches 42; Indels 1; Gaps 1;  
QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQ 59  
DB 518 GPDGMPGPPGPTGTSIAHGFLTRHSQTTDAPQPHGTVQIYEGSFLYVQGNORAHQ 577  
QY 60 DLGTGSLQRTFTMPFLFCNVNDVNCNPNFASNDYSLWSTPALMPNNAPITGRALEPYI 119  
DB 578 DLGTAGSCLRRFSTMPFECNINNVNPNFASNDYSLWSTPEPMPMSQPLKQSGIOFFI 637  
QY 120 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCL 179  
DB 638 SRCVACEAPAVVIAVHSQTIQIHPCPQGWDSLWATGYSEFMHTSAGSEGTGQALASPGSCL 697  
QY 180 EFRASPFLECHGRGTCNYNSYSFWLASLNPERMPKPIPTVTKAGELEKIISRC 236  
DB 698 EFRSAPFIECHGRGTCNYNSYSFWLATVDVSMFSKQSETLKAGDLRTRISRC 754

RESULT 10  
B49736  
collagen alpha 3(IV) chain, medium splice form - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 12-Nov-1999 #text\_change 17-Mar-2000  
C:Accession: B49736; D49736; S69111  
R.Feng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A:Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.  
A:Reference number: A49736; MUID:94124597; PMID:8294492  
A:Accession: B49736  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 169-220 <FEN1>  
A:Accession: D49736  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 22-220 <FEN2>  
A:Cross-references: GB:U02519; NID:9409106; PIDN:AAA18942.1; PID:9409107  
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank  
R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; W.  
Eur. J. Biochem. 229, 754-760, 1995  
A:Title: Characterization and expression of multiple alternatively spliced transcripts  
A:Reference number: S69111; MUID:95278230; PMID:7758473  
A:Accession: S69111  
A:Molecule type: mRNA  
A:Residues: 1-45,169-204,'L',206-220 <PEN>  
C:Comment: For the complete sequence of the long splice form, see PIR:CGHUB.  
C:Genetics:  
A:Gene: GDB:COL4A3  
A:Cross-references: GDB:128351; OMIM:120070  
A:Map position: 2q36-2q37  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extract  
F:1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status prec  
F:1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status  
F:22-220/Dominant: carboxyl-terminal nonhelical, NCI <NCL>  
F:234-134/Dominant: collagen IV carboxyl-terminal repeat <CTL>

Query Match 65.1%; Score 872; DB 2; Length 220;  
Best Local Similarity 99.4%; Pred. No. 1.3e-71;  
Matches 158; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQ 60  
DB 10 GLKGRGDSGSPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNORAHQ 69



QY 61 LGTGLSCLORTTTPFPFCNVNDVNCNPFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 120  
DB 70 LGTGLSCLORTTTPFPFCNVNDVNCNPFASRNDYSYWLSTPALMPMNPAPITGRALPEYIS 129  
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTIM 159  
DB 130 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTIM 168  
RESULT 11  
S40991  
collagen alpha 1(IV) chain precursor - Caenorhabditis elegans  
N:Alternate names: protein K04H4.1  
C:Species: Caenorhabditis elegans  
C:Date: 03-May-1994 #sequence\_revision 02-Aug-1994 #text\_change 13-Aug-1999  
R:Ainscough, R.  
C:Accession: S40991; S44442; S13651; B34476  
submitted to the EMBL Data Library, October 1993  
A:Reference number: S40991  
A:Accession: S40991  
A:Molecule type: DNA  
A:Residues: 1-1744 <AIN>  
A:Cross-references: EMBL:X56978; NID:G414627; PID:G414628  
R:Kramer, J.M.  
submitted to the EMBL Data Library, December 1990  
A:Reference number: S44442  
A:Accession: S44442  
A:Molecule type: DNA  
A:Residues: 1-129, 'GFGCMPLAGPQSGONGNPGRLGSLGPPGEGVNSQGRKGKVGSGSGVGPLP', 209-281, 'PV  
15, 'D', 817-1260, 'P', 1262-1707, 'P', 1709-1744 <KRA>  
A:Cross-references: EMBL:X56979; NID:G6675; PIDN:CAA40299.1; PID:G6676  
R:Guo, X.; Johnson, J.J.; Kramer, J.M.  
Nature 349, 707-709, 1991  
A:Title: Embryonic lethality caused by mutations in basement membrane collagen of C. ele  
A:Reference number: S13651; NID:G1141582; PMID:1996137  
A:Accession: S13651  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-129, 'GFGCMPLAGPQSGONGNPGRLGSLGPPGEGVNSQGRKGKVGSGSGVGPLP', 209-281, 'PV  
15, 'D', 817-1260, 'P', 1262-1515 <GUL>  
A:Cross-references: EMBL:X56979  
R:Guo, X.; Kramer, J.M.  
J. Biol. Chem. 264, 17574-17582, 1989  
A:Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1  
A:Reference number: A34476; NID:90008929; PMID:2793871  
A:Accession: B34476  
A:Molecule type: DNA  
A:Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>  
A:Cross-references: EMBL:X05067; NID:G156255; PIDN:AAB59179.1; PID:G156256  
C:Genetics:  
A:Gene: clb-2, emb-9  
A:Map position: 3  
A:Interons: 23/2; 79/1; 152/2; 289/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F:43-1515/Domain: collagenous, triple helix #status predicted <COL>  
F:93-95/Region: cell attachment (R-G-D) motif  
F:1053-1055/Region: cell attachment (R-G-D) motif  
F:1396-1398/Region: cell attachment (R-G-D) motif  
F:1516-1744/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>  
F:1516-1627, 1628-1744/Region: duplication  
F:1580-1586, 1691-1697/Disulfide bonds: #status predicted  
Query Match 63.6%; Score 852.5; DB 2; Length 1744;  
Best Local Similarity 60.7%; Pred. No. 7.2e-69;  
Matches 148; Conservative 35; Mismatches 60; Indels 1; Gaps 1;  
QY 1 GLKXKXGDSGSPATWT-TRGFVTRHSQTTAIPSPCEGTVPLYSQFSLFVQGNORAHGQ 59  
DB 1501 GLPPTGYPGSGCGWAPSGGF-FAKHSQTTAVPQPPGASQLWEGYSLLYVQGNRAGSQ 1560  
QY 60 DLGTLGSCLORTTTPFPFCNVNDVNCNPFASRNDYSYWLSTPALMPMNPAPITGRALPEYI 119

DB 1561 DLGCGSCLSKENTPFPCNWNVSCHVSSRNDYSFWLSTDEPMTPMNPVTGTAIRPYI 1620  
QY 120 SRTCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTIMTSAGSEGTGQALASPGSCL 179  
DB 1621 SRCVACEVPTQIIAIVHSQDTSVPQCPQGWGNTGYSFVMTAAAGETGOSLOSBSCL 1680  
QY 180 EBFRAFPFLECHGRGTCNVYNSYSFWLASLNPFRMFKPIPTSTVKAGELSKIIISRCQVC 239  
DB 1681 EBFRAVPFIECHGRGTCNYATNHGFWLSIVDQDKFKPKMSQTLKAGGLKDRVSRQVC 1740  
QY 240 MKKR 243  
DB 1741 LKNR 1744  
RESULT 12  
S49488  
collagen alpha 3(IV) chain - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: S49488  
R:Oberbauer, I.  
submitted to the EMBL Data Library, October 1994  
A:Description: Cloning of the NCI domains fo the minor collagen IV chains of mouse via  
ells.  
A:Reference number: S49487  
A:Accession: S49488  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-161 <OBE>  
A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAAS7689.1; PID:G559916  
C:Superfamily: collagen alpha 1(IV) chain  
Query Match 62.2%; Score 834; DB 2; Length 161;  
Best Local Similarity 92.5%; Pred. No. 2.6e-68;  
Matches 149; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
QY 66 SCLQRTTTPFPFCNVNDVNCNPFASRNDYSYWLSTPALMPMNPAPITGRALPEYISRCTVC 125  
DB 1 SCLQRTTTPFPFCNINNVNCFASRNDYSYWLSTPALMPMDNPISGRALPEYISRCTVC 60  
QY 126 EGPFAIAIAVHSQTTDIPPCPHGWISLWKGFSTIMTSAGSEGTGQALASPGSCLBEFRAS 185  
DB 61 EGPFAIAIAVHSQTTAIPPCPDWVSLWKGFSTIMTSAGSEGTGQALASPGSCLBEFRAS 120  
QY 186 PFLECHGRGTCNVYNSYSFWLASLNPFRMFKPIPTSTVKA 226  
DB 121 PFIECHGRGTCNVYNSYSFWLASLNPFRMFKPIPTSTVKA 161  
RESULT 13  
A45407  
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C:Accession: A45407; A43903; A23940  
R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.  
J. Biol. Chem. 268, 5249-5254, 1993  
A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and an  
A:Reference number: A45407; NID:93186842; PMID:8444899  
A:Accession: A45407  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1752 <2XP>  
A:Note: sequence extracted from NCBI backbone (NCBIP:126841)  
R:Wessel, G.M.; Etkin, M.; Benson, S.  
Dev. Biol. 148, 261-272, 1991  
A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously proc  
A:Reference number: A43903; NID:92038439; PMID:1936564  
A:Accession: A43903  
A:Status: preliminary  
A:Molecule type: mRNA

A;Residues: 'P', 633-1537, 'G' <WES>  
A;Cross-references: GB:S64572; NID:G238616; PIDN:AA820270.1; PID:G238617  
A;Note: Sequence extracted from NCBI Backbone (NCBIN:64572, NCBIP:64573)  
R;Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986  
A;Title: Structure and developmentally regulated expression of a Strongylocentrotus purp  
A;Reference number: A23940; MUID:86205894; PMID:3459186  
A;Accession: A23940  
A;Molecule type: DNA  
A;Residues: 742-812 <VEN>  
A;Cross-references: ENBL:M13206  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
F:29-161/Domain: amino-terminal nonhelical, 7S <7SD>  
F:162-1523/Domain: interrupted helical  
F:1524-1752/Domain: carboxyl-terminal nonhelical, NCl <NCl>  
F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT>  
F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT>  
F:125/Modified site: alllysine (Lys) #status predicted

Query Match 62.1%; Score 832; DB 2; Length 1752;  
Best Local Similarity 59.3%; Pred No. 5.2e-67;  
Matches 144; Conservative 37; Mismatches 60; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
DB 1512 GPGGTKEAGIPG--SSSGFTTRHSQTTSIPQCPQGCTAKMWHGYSLLFVQNERGHQD 1569

QY 61 LGTGLSCLORETTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMMAPIITGRALEPYIS 120  
DB 1570 LGKPGSCUKRSTPFPFLFCNNVCHVASRNDYSYWLSTTEPPMNMMAPIITRGGLQOPFIS 1629

QY 121 RCTVCEGPAIAVHVSQTTDPPCPHGWSLWKGSFTMFTSAGSEGTGQALASPGSCLE 180  
DB 1630 RCWCEAPQVLTVHSQTVNPDCPDGVLWIGYSFWHTGPGEGSQMLSSPGSCLE 1689

QY 181 EFRASPFLCEHGRGTCNYSNSYSFWLASLNPFRMFKPIPTSTVKAGLEKIIISRCQVCM 240  
DB 1690 DFRSPFTECHGDGKCNYYATYTFWFSSITGNACFTMPQSETLKAGSLRTRVSRCAVCL 1749

QY 241 KKR 243  
DB 1750 RNQ 1752

RESULT 14  
S16366  
collagen alpha 2(IV) chain precursor - pig roundworm  
C;Species: Ascaris suum (pig roundworm)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: S16366  
R;Pettitt, J.; Kingston, I.B.  
J. Biol. Chem. 266, 16149-16156, 1991  
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti  
A;Reference number: S16366; MUID:91340768; PMID:1714907  
A;Accession: S16366  
A;Molecule type: mRNA  
A;Residues: 1-1763 <JBI>  
A;Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649  
C;Genetics:  
A;Introns: 229/3; 305/3; 360/3; 424/1; 489/1; 548/3; 790/1; 891/1; 963/1;  
A;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid  
F:1-26/Domain: signal sequence #status predicted <Sig>  
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>  
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>  
F:43-1529/Domain: collagenous #status predicted <COL>  
F:197-199/Region: cell attachment (R-G-D) motif  
F:1530-1763/Domain: carboxyl-terminal nonhelical, NCl #status predicted <NCl>  
F:1530-1638/Domain: repeat NCl #status predicted <NCl1>  
F:1639-1763/Domain: repeat NCl #status predicted <NCl2>  
F:31.34.39.41.536.539/Disulfide bonds: interchain #status predicted  
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1593-1599, 1702-1709/Disulfide bonds: #status predicted

Query Match 59.4%; Score 796.5; DB 2; Length 1763;  
Best Local Similarity 57.9%; Pred. No. 8.7e-64;  
Matches 140; Conservative 41; Mismatches 58; Indels 3; Gaps 2;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
DB 1515 GLPGSGPPGPPPSYKDGFLLVKHSQISEVPQCPGVMKLVWDGYSLLYIEGNEKSHNQD 1574

QY 61 LGTGLSCLORETTPPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMMAPIITGRALEPYIS 120  
DB 1575 LGHAGSCLSRFTSTPFLFCNVNVCNYSRNDKSYWLSTTA--PIPMVPSVSEGGTEPYIS 1632

QY 121 RCTVCEGPAIAVHVSQTTDPPCPHGWSLWKGSFTMFTSAGSEGTGQALASPGSCLE 180  
DB 1633 RCACVEAPANVAVHSQTIQIPNCPCGNWSLWIGYSFAMHTGAGAGGGQSLSSPGSCLE 1692

QY 181 EFRASPFLCEHGRGTCNYSNSYSFWLASLNPFRMFKPIPTSTVKAGLEKIIISRCQVC 239  
DB 1693 DFRATPFTECNARGCTCHVFANKFSFWLTITIEDDQCPRIPESETLKAGSLRTRVSRQVC 1752

QY 240 MK 241  
DB 1753 IR 1754

RESULT 15  
T29350  
hypothetical protein F01G12.5a - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T29350  
R;Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A;Description: The sequence of C. elegans cosmid F01G12.  
A;Reference number: Z20611  
A;Accession: T29350  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1758 <WUX>  
A;Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a  
A;Experimental source: strain Bristol N2; clone F01G12  
C;Genetics:  
A;Gene: CESP:F01G12.5a  
A;Map position: X  
A;Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3;  
C;Superfamily: collagen alpha 1(IV) chain

Query Match 58.5%; Score 783.5; DB 2; Length 1758;  
Best Local Similarity 55.3%; Pred. No. 1.3e-62;  
Matches 142; Conservative 41; Mismatches 59; Indels 15; Gaps 4;

QY 1 GLKGRGDSGP-----ATWTRGFVTRHSQTATPSCPEGTVPVLYSGFSLFVQ 51  
DB 1504 GLDGQPGGPGAGLPGAGAGPARDGDFVLVKSQTTEVRCPEGQTKLWDGYSLLYIE 1563

QY 52 GNORAHQDGLTGLSGCLQRTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNMMAPIIT 111  
DB 1564 GNEKSHNQDLGHAGSCLQRTSTPFLFCDFNNVCNYSRNDKSYWLSTSEAIIP--MMPVN 1621

QY 112 GRALPYSRCTVCEGPAIAVHVSQTTDIPCPHGWSLWKGSFTMFTSAGSEGTGQA 171  
DB 1622 EREIPYSRCAVCAPANTAVHSQTIQIPNCPCGNWSLWIGYSFAMHTGAGAGGGQGS 1681

QY 172 LASPGSCLEEFSPFLCEHGRGTCNYSNSYSFWLASLNPFRMFKPIPTSTVKAGLE 230  
DB 1682 LSSPGSCLEDFRATPFIECNARGSGCHVFANKFSFWLTITIDNDSFKVPSQETLKSGNLR 1741

QY 231 KIISRCQVCMKK---RH 244  
DB 1742 TRVSRQVCVKSTDGRH 1758



Mon Apr 5 07:53:05 2004

us-10-032-221b-10.rpr

Page 10

Search completed: April 5, 2004, 07:05:33  
Job time : 51.2179 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 30.7215 Seconds  
(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKGRGDSGPATWTRGP.....KAGELEKIISRCQVCMKKEH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1340	100.0	1670	1 CA34 HUMAN	Q01955 homo sapien
2	1210.5	90.3	471	1 CA34 BOVIN	Q28084 bos taurus
3	960	71.6	1669	1 CA14 HUMAN	P02462 mus musculu
4	951	71.0	1669	1 CA14 MOUSE	P02463 mus musculu
5	943.5	70.4	1685	1 CA54 HUMAN	P29400 homo sapien
6	922.5	68.8	754	1 CA54 CANFA	Q28247 canis famil
7	845.5	63.1	1758	1 CA14 CAEEL	P17139 caenorhabdi
8	795.5	59.4	1763	1 CA24 ASGSU	P27393 ascaris suu
9	772.5	57.6	1758	1 CA24 CAEEL	P17140 caenorhabdi
10	760.5	56.8	1707	1 CA24 MOUSE	P08122 mus musculu
11	755.5	56.4	1712	1 CA24 HUMAN	P08572 homo sapien
12	739	55.1	1691	1 CA64 HUMAN	Q14031 homo sapien
13	709	52.9	453	1 CA44 BOVIN	Q29442 bos taurus
14	702	52.4	623	1 CA44 RABIT	P55787 cryptolagus
15	696	51.9	1690	1 CA44 HUMAN	P33420 homo sapien
16	694.5	51.8	1775	1 CA14 DROME	P08120 drosophila
17	98.5	7.2	437	1 COX1 SCHPO	P07657 schizosacch
18	96.5	7.2	4391	1 PGEM HUMAN	P38160 homo sapien
19	89.5	6.7	3707	1 PGEM MOUSE	Q05793 mus musculu
20	86.5	6.5	755	1 MTS1 HUMAN	O43312 homo sapien
21	81.5	6.1	539	1 IL2B MOUSE	P16297 mus musculu
22	80	6.0	599	1 G363 LEICH	P15706 leishmania
23	79.5	5.9	1597	1 M3K4 MOUSE	Q08648 mus musculu
24	79	5.9	837	1 GCSR MOUSE	P40223 mus musculu
25	79	5.9	1877	1 PK55 MOUSE	Q04592 mus musculu
26	78.5	5.9	442	1 CHMO AMATR	Q03101 anarantus
27	78	5.8	669	1 MYBE AVILE	P01105 avian leuko
28	77.5	5.8	1202	1 JAG2 RAT	P27607 rattus norv
29	77.5	5.8	1328	1 AGR1 DISOM	Q04044 discopyge o
30	77.5	5.8	1391	1 N155 HUMAN	O75694 homo sapien
31	77	5.7	366	1 CAS4 EPHMC	P18503 ephydattia m
32	77	5.7	369	1 DNAB NITEU	O66431 nitrosomona
33	77	5.7	1639	1 LMGI1 DROME	P15215 drosophila

34 76.5 5.7 155 1 HOPD\_SALTY  
35 76.5 5.7 680 1 CALA\_HUMAN  
36 76.5 5.7 770 1 PKC7\_MOUSE  
37 76.5 5.7 783 1 PKC7\_RAT  
38 76.5 5.7 5703 1 MUSB\_HUMAN  
39 76 76 379 1 DNAB\_LEGPN  
40 76 76 617 1 SPH2\_MOUSE  
41 76 76 633 1 MUTL\_PSEAE  
42 76 76 674 1 CALA\_CHICK  
43 75.5 5.6 785 1 PKC7\_HUMAN  
44 75.5 5.6 1411 1 Y297\_HUMAN  
45 75.5 5.6 1959 1 AGRI\_RAT

#### ALIGNMENTS

RESULT 1  
CA34\_HUMAN  
ID CA34\_HUMAN STANDARD; PRT; 1670 AA.  
AC Q01955; Q9BQT2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94364394; PubMed=8083201;  
RA Mariyana M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RL J. Biol. Chem. 269:23013-23017 (1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;  
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND  
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;  
PRO-574; GLU-1269 AND PRO-1474.  
RX MEDLINE=21064596; PubMed=1134255;  
RA Heidet L., Arrondelet C., Forestier L., Cohen-Solal L., Mollet G.,  
Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106 (2001).  
RN [4]  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:19780-19784 (1992).  
RN [5]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554 (1991).  
RN [6]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RX TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;



Query Match 100.0%; Score 1340; DB 1; Length 1670;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-116;  
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKKGKDGSSPATWTRGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60  
 DB 1427 GLKKGKDGSSPATWTRGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 1486

QY 61 LGTLGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAIPITGRALPEPYIS 120  
 DB 1487 LGTLGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAIPITGRALPEPYIS 1546

QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSGSGTGOALASPGSCLE 180  
 DB 1547 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSGSGTGOALASPGSCLE 1606

QY 181 EFRASPLECHGRCTCNYSNSYFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCM 240  
 DB 1607 EFRASPLECHGRCTCNYSNSYFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCM 1666

QY 241 KGRH 244  
 DB 1667 KGRH 1670

## RESULT 2

CA34\_BOVIN STANDARD; PRT; 471 AA.  
 ID CA34\_BOVIN  
 AC Q28084;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Collagen alpha 3(IV) chain (Fragment).  
 GN COL4A3.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lens;  
 RX MEDLINE=91093146; PubMed=1985905;  
 RA Morrison K.E., Germino G.G., Reeder S.T.;  
 RT "Use of the polymerase chain reaction to clone and sequence a cDNA  
 encoding the bovine alpha 3 chain of type IV collagen.";  
 RL J. Biol. Chem. 266:34-39(1991).  
 RN [2]  
 RP SEQUENCE OF 227-258.  
 RC TISSUE=Kidney;  
 RX MEDLINE=90202779; PubMed=2318822;  
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;  
 RT "Glomerular basement membrane. Identification of a fourth chain,  
 alpha 4, of type IV collagen.";  
 RL J. Biol. Chem. 265:5466-5469(1990).  
 RN [3]  
 RP SEQUENCE OF 227-254.  
 RX MEDLINE=89330844; PubMed=3417661;  
 RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;  
 RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain  
 of collagen IV.";  
 RL J. Biol. Chem. 263:13374-13380(1988).  
 RN [4]  
 RP SEQUENCE OF 227-244.  
 RX MEDLINE=87222419; PubMed=2438283;  
 RA Buckowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,  
 RA Hudson B.G.;  
 RT "Localization of the Goodpasture epitope to a novel chain of basement  
 membrane collagen.";  
 RL J. Biol. Chem. 262:7874-7877(1987).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 glomerular basement membranes (GBM), forming a 'chicken-wire',

meshwork together with laminins, proteoglycans and entactin/  
 nidogen.  
 -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
 alpha 6(IV), each of which can form a triple helix structure  
 with 2 other chains to generate type IV collagen network.  
 -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
 -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 domain (NC1) at their C-terminus, frequent interruptions of the  
 G-X-Y repeats in the long central triple-helical domain (which may  
 cause flexibility in the triple helix), and a short N-terminal  
 triple-helical 7S domain.  
 -!- PTM: Prolines at the third position of the tripeptide repeating  
 unit (G-X-Y) are hydroxylated in some or all of the chains.  
 -!- PTM: Type IV collagens contain numerous cysteine residues which  
 are involved in inter- and intramolecular disulfide bonding. 12 of  
 these, located in the NC1 domain, are conserved in all known type  
 IV collagens.  
 -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----

EMBL; M63139; AB062708.1; -;  
 FIR; A39024; A39024.  
 DR InterPro; IPR008150; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 4.  
 DR ProDom; PD003323; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 238  
 FT SITE 239 471  
 FT SITE 106 108  
 FT MOD\_RES 232 232  
 FT MOD\_RES 238 238  
 FT DISULFID 261 352  
 FT DISULFID 294 349  
 FT DISULFID 306 312  
 FT DISULFID 371 466  
 FT DISULFID 405 463  
 FT DISULFID 417 423  
 FT CONFLICT 253 253  
 FT CONFLICT 471 AA; 47595 MW; C03B6F14E7008DE CRC64;  
 SQ SEQUENCE 471 AA; 47595 MW; C03B6F14E7008DE CRC64;  
 Query Match 90.3%; Score 1210.5; DB 1; Length 471;  
 Best Local Similarity 90.6%; Pred. No. 3.7e-105;  
 Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKKGKDGSSPATWTT-RGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQ 59  
 DB 227 GLKKGKDGSSPATWTT-RGVFTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQ 286

QY 60 DLGTGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAIPITGRALPEYI 119  
 DB 287 DLGTGSCLOQFTTTPFLFCNVNDVCFNPNFASRNDYSYWLSTPALPMNNAIPITGRALPEYI 346

QY 120 SRTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSGSGTGOALASPGSCLE 179  
 DB 347 SRTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSGSGTGOALASPGSCLE 406

QY 180 EFRASPLECHGRCTCNYSNSYFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCM 239  
 DB 407 EFRASPLECHGRCTCNYSNSYFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCM 466

QY 240 MKKR 243

Db 467 MMR 470

|||

RESULT 3

CA14 HUMAN

ID -CA14 HUMAN STANDARD; PRT; 1669 AA.

AC P02462;

DT 21-JUN-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 1(IV) chain precursor.

GN COL4A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89340433; PubMed=2701944;

RA Soinenen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;

RT "Structural organization of the gene for the alpha 1 chain of human type IV collagen."

RL J. Biol. Chem. 264:13565-13571(1989).

[2]

RP SEQUENCE OF 46-1257 FROM N.A.

RX TISSUE=Placenta;

RX MEDLINE=88083584; PubMed=3691802;

RA Soinenen R., Haka-Risku T., Prockop D.J., Tryggvason K.;

RT "Complete primary structure of the alpha 1-chain of human basement membrane (type IV) collagen."

RL FEBS Lett. 225:188-194(1987).

[3]

RP SEQUENCE OF 1-943 FROM N.A.

RX TISSUE=Placenta;

RX MEDLINE=88029471; PubMed=3311751;

RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W., Deutzmann R., Kuehn K.;

RT "Completion of the amino acid sequence of the alpha 1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain."

RL Eur. J. Biochem. 168:529-536(1987).

[4]

RP SEQUENCE OF 28-243.

RX MEDLINE=86004708; PubMed=4043082;

RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;

RT "Amino acid sequence of the N-terminal aggregation and cross-linking region (7S domain) of the alpha 1 (IV) chain of human basement membrane collagen."

RL Eur. J. Biochem. 152:213-219(1985).

[5]

RP SEQUENCE OF 534-1447.

RX MEDLINE=8503629; PubMed=6434307;

RA Babel W., Glanville R.W.;

RT "Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sequence of a 914-residue-long pepsin fragment from the alpha 1(IV) chain."

RL Eur. J. Biochem. 143:545-556(1984).

[6]

RP SEQUENCE OF 1256-1669 FROM N.A.

RX MEDLINE=85207819; PubMed=2581969;

RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R., Cheung M.-C., Prockop D.J., Boyd C.D.;

RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV procollagen reveal an unusual homology of amino acid sequences in two halves of the carboxyl-terminal domain."

RL J. Biol. Chem. 260:7681-7687(1985).

[7]

RP SEQUENCE OF 1259-1669 FROM N.A.

RX MEDLINE=85216555; PubMed=2582422;

RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J., Keralides N.A., Myers J.C.;

RT "Restricted homology between human alpha 1 type IV and other procollagen chains.";

Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).

[8]

RP SEQUENCE OF 1-28 FROM N.A.

RX MEDLINE=89034231; PubMed=3182844;

RA Soinenen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;

RT "The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region."

RL J. Biol. Chem. 263:17217-17220(1988).

[9]

RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.

RX TISSUE=Placenta;

RX MEDLINE=89005112; PubMed=2844531;

RA Siebold B., Deutzmann R., Kuehn K.;

RT "The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen."

RL Eur. J. Biochem. 176:617-624(1988).

CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 5(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Lysines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M26576; AAA53098.1; JOINED.

EMBL; J04217; AAA53098.1; JOINED.

EMBL; M26550; AAA53098.1; JOINED.

EMBL; M26540; AAA53098.1; JOINED.

EMBL; M26542; AAA53098.1; JOINED.

EMBL; M26543; AAA53098.1; JOINED.

EMBL; M26544; AAA53098.1; JOINED.

EMBL; M26545; AAA53098.1; JOINED.

EMBL; M26546; AAA53098.1; JOINED.

EMBL; M26547; AAA53098.1; JOINED.

EMBL; M26537; AAA53098.1; JOINED.

EMBL; M26538; AAA53098.1; JOINED.

EMBL; M26548; AAA53098.1; JOINED.

EMBL; M26549; AAA53098.1; JOINED.

EMBL; M26551; AAA53098.1; JOINED.

EMBL; M26552; AAA53098.1; JOINED.

EMBL; M26553; AAA53098.1; JOINED.

EMBL; M26554; AAA53098.1; JOINED.

EMBL; M26555; AAA53098.1; JOINED.

EMBL; M26556; AAA53098.1; JOINED.

EMBL; M26557; AAA53098.1; JOINED.

EMBL; M26559; AAA53098.1; JOINED.

EMBL; M26558; AAA53098.1; JOINED.

EMBL; M26559; AAA53098.1; JOINED.

EMBL; M26560; AAA53098.1; JOINED.



```

RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198625;
RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RR SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL J. Biol. Chem. 263:8706-8709(1988).
RN [9]
RR SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; J03758; AAA37439.1; -
DR EMBL; M23333; AAA51625.1; -
DR EMBL; J04694; AAA50292.1; -
DR EMBL; X06777; CAA28946.1; -
DR EMBL; X02201; CAA26132.1; -
DR EMBL; M15832; AAA37340.1; -
DR EMBL; M14042; AAA37342.1; -
DR EMBL; M12879; AAA37343.1; -
DR EMBL; M13024; -; NOT ANNOTATED_CDS.
DR EMBL; M13025; -; NOT ANNOTATED_CDS.
DR EMBL; M13026; AAA37344.1; -
DR EMBL; M13027; AAA37345.1; -
DR EMBL; M13043; AAA37346.1; -
DR EMBL; J04448; AAA37437.1; -
DR PIR; A33525; CGMS48.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; ProCollagnc4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.

```

```

FT SIGNAL 1 27
FT PROPEP 28 172
FT CHAIN 173 1669
FT DOMAIN 173 1440
FT DOMAIN 1441 1669
FT DISULFID 1460 1551
FT DISULFID 1493 1548
FT DISULFID 1505 1511
FT DISULFID 1570 1665
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CARBOHYD 126 126
FT CONFLICT 26 26
FT CONFLICT 186 186
FT CONFLICT 319 319
FT CONFLICT 369 369
FT CONFLICT 403 403
FT CONFLICT 481 481
FT CONFLICT 493 493
FT CONFLICT 712 712
FT CONFLICT 813 813
FT CONFLICT 982 982
FT CONFLICT 1397 1397
FT SEQUENCE 1369 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 71.0%; Score 951; DB 1; Length 1669;
Best Local Similarity 68.6%; Pred. No. 2.3e-80;
Matches 166; Conservative 34; Mismatches 40; Indels 2; Gaps 1;

QY 1 GLKRGDSGSPATWTRGTFVTRHSQTTPSPQECTVPLYSQSFVQGNORAHQGD 60
DB 1429 GLPSMGPPGTPS--VDHGLFVTRHSQTDDPLCFPGTKLYHGYSLLYVQGNRAHQGD 1486
QY 61 LGTGLSCLORETFMPFFFCNVNDVNCVFNASNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 1487 LGTAGSCLRKSTMPFFFCNVNDVNCVFNASNDYSYWLSTPALPMNMAPITGRALEPYIS 1546
QY 121 RCTVCEGPAIAIVHSQTTPPCPHGWI SIWKGFSTFIMFSTAGSGTGALASPGSCLE 180
DB 1547 RCACVEAPAMVAVHSQTTPPCPHGWI SIWKGFSTFIMFSTAGSGTGALASPGSCLE 1606
QY 181 EFRASPFLCHGRGTCNYSYNSYSYFWLASLNPFRKPIPTVKAGELKFIISRCQVCM 240
DB 1607 EFRSAPFTECHGRGTCNYSYNSYSYFWLASLNPFRKPIPTVKAGELRTHVSRQVCM 1666
QY 241 KK 242
DB 1667 RR 1668

RESULT 5
CA54_HUMAN
ID CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5(IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CVS-521.
RC TISSUE=Kidney;
RX MEDLINE=92316923; PubMed=1352287;

```

RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen  
RT chain and identification of a single-base mutation in exon 23  
RT converting glycine 541 in the collagenous domain to cysteine in an  
RT Alport syndrome patient.";  
RL J. Biol. Chem. 267:12475-12481(1992).  
RN [3]  
RP SEQUENCE OF 85-1685 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=90337990; PubMed=2380186;  
RA Pihlajaniemi T., Pohjola-Evonen E.R., Myers J.C.;  
RT "Complete primary structure of the triple-helical region and the  
RT carboxyl-terminal domain of a new type IV collagen chain, alpha  
RT 5(IV).";  
RL J. Biol. Chem. 265:13758-13766(1990).  
RN [4]  
RP SEQUENCE OF 924-1685 FROM N.A.  
RX MEDLINE=91169491; PubMed=2004755;  
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
RT "Characterization of the 3' half of the human type IV collagen alpha  
RT 5 gene that is affected in the Alport syndrome.";  
RL Genomics 9:1-9(1991).  
RN [5]  
RP SEQUENCE OF 914-1685 FROM N.A.  
RX MEDLINE=90160375; PubMed=1889491;  
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,  
RA Tryggvason K.;  
RT "Identification of a distinct type IV collagen alpha chain with  
RT restricted kidney distribution and assignment of its gene to the  
RT locus of X chromosome-linked Alport syndrome.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RP SEQUENCE OF 1442-1471 FROM N.A.  
RX MEDLINE=90252791; PubMed=2339699;  
RA Myers J.C., Jones T.A., Pohjola-Evonen E.R., Kadri A.S., Goddard A.D.,  
RA Sheer D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
RT to the region of the X chromosome containing the Alport syndrome  
RT locus.";  
RL Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RP SEQUENCE OF 1-20 FROM N.A.  
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
RA Marynen P.;  
RT Submitted (SEP-1994) to the ENBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).  
RX MEDLINE=94133540; PubMed=8301933;  
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
RA Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
RT cells: a complex mutation in the COL4A5 gene of an Alport patient  
RT deletes the NCI domain.";  
RL Kidney Int. 44:1316-1321(1993).  
RN [9]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
RL Hum. Mutat. 9:477-499(1997).  
RN [10]  
RP VARIANT AS SER-1564.  
RX MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
RA Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome.";  
RL Genomics 9:10-18(1991).  
RN [11]  
RP VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
RA Tryggvason K., Gubler M.-C., Antignac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
RT (IV) chain associated with X-linked Alport syndrome: characterization  
RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
RT cDNA fragments.";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [12]  
RP VARIANT AS GLU-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
RA Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome.";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [13]  
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=8406498;  
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
RA Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
RA van Oost B.A., Smeets H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of  
RT patients with Alport syndrome.";  
RL Genomics 17:485-489(1993).  
RN [14]  
RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;  
RX ARG-869; ARG-872 AND CYS-1241.  
RA MEDLINE=95322976; PubMed=7599631;  
RX Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
RT "Detection of 12 novel mutations in the collagenous domain of the  
RT COL4A5 gene in Alport syndrome patients.";  
RL Hum. Mutat. 5:197-204(1995).  
RN [15]  
RP VARIANT AS ARG-1649.  
RX MEDLINE=96213750; PubMed=8651292;  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States.";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [16]  
RP VARIANTS AS.  
RX MEDLINE=96213754; PubMed=8651296;  
RA Renieri A., Bruttini N., Galli L., Zarelli P., Neri T.M., Rossetti S.,  
RA Turco A.E., Heiskari M., Zhou J., Gusmano R., Massella L., Banfi G.,  
RA Scolari F., Seesa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
RA Savi M., Ballabio A., de Marchi M.;  
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
RT exons of the COL4A5 gene.";  
RL Am. J. Hum. Genet. 58:1192-1204(1996).  
RN [17]  
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; ASN-664 AND  
RP MET-1428.  
RX MEDLINE=97094179; PubMed=8940267;  
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,  
RA Glatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
RA Gubler M.-C., Antignac C.;  
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
RT syndrome.";  
RL Am. J. Hum. Genet. 59:1221-1232(1996).  
RN [18]  
RP VARIANT AS ASP-1498.  
RX MEDLINE=96233932; PubMed=8829632;  
RA Tverskaya S., Bobrymina V., Tsalykova F., Ignatova M.,  
RA Krasnopol'skaya X., Evgrafov O.;  
RT "Substitution of Ala98D in noncollagen domain of alpha 5(IV) collagen  
RT chain associated with adult-onset X-linked Alport syndrome.";  
RL Hum. Mutat. 7:149-150(1996).  
RN [19]  
RP VARIANT AS GLN-1677.  
RX MEDLINE=97295089; PubMed=9150741;  
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
RT "Common ancestry of three Ashkenazi-American families with Alport  
RT syndrome and COL4A5 R1677Q.";  
RL Hum. Genet. 99:681-684(1997).



[20]  
RN VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
RP AND ASP-1596.  
RX MEDLINE=98112435; PubMed=9452056;  
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
RA Pignatti G.F., Galli L., Pruttini M., Renieri A., Mingarelli R.,  
RA Trivelli A., Pinciaroli A.R., Ragatolo M., Rizzoni G.F., de Marchi M.,  
RA "Missense mutations in the COL4A5 gene in patients with X-linked  
RT Alport syndrome.";  
RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
[21]  
RN VARIANTS AS VAL-420; 456-PRO-PRO-ARG-458 DEL; ASP-573; ASP-624; ASP-635;  
RP 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
RX MEDLINE=98112435; PubMed=9452056;  
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tunelius T., Hertz J.M.,  
RA Barker D.F., Gregory M.C., Atkin C.B., Stykardottir U., Neumann H.,  
RA Springate J., Shows T.B., Petterson E., Tryggsavson K.,  
RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
RT Alport syndrome using PCR and direct DNA sequencing.";  
EL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
[22]  
RN VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
RX MEDLINE=20030197; PubMed=10561141;  
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.,  
RT "Detection of mutations in the COL4A5 gene in over 90% of male  
RT patients with X-linked Alport's syndrome by RFLP-PCR and direct  
RT sequencing.";  
RL Am. J. Kidney Dis. 34:854-862(1999).  
[23]  
RN VARIANT AS ARG-822.

Query Match 70.4%; Score 943.5; DB 1; Length 1685;  
Best Local Similarity 67.9%; Pred. No. 1.2e-79;  
Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;  
QY 1 GLKGRKDGSGSPATWT-TRGFVFTRHSSQTTPAIPSCPGTFLVSGFSLVQGNRAHQ 59  
DB 1442 GPDGLQGGPPGPGTSSVAHGFLLTRHSQTTPAIPSCPGTFLVSGFSLVQGNRAHQ 1501  
QY 60 DLGTLGSLQFTTTPFLFCNVNDVCFNFSRNDYSYMLSTPALPMNMAITGEALPYI 119  
DB 1502 DLGTAGSLRFRSTPMFNCNINNVCFNFSRNDYSYMLSTPEPMNSQPLKGGSTQFPI 1561  
QY 120 SRCTVCEGPAIAVHSOTTDIPCPHGWISLWGFSPIMFTSAGSGTGOALASPGSCL 179  
DB 1562 SRCVCEAPAVVIAVHSQTIQIPHCPCQWDSLWIGSYFMHTSAGSGGOALASPGSCL 1621  
QY 180 EFRASPLECHGRGTCNYNSYSFVLASLNPFRMFKPIPTSVKAGELEKILSRQVC 239  
DB 1622 EFRSAPFTECHGRGTCNYNSYSFVLATVDVDMFSKQSETKAGDLNTRISRCQVC 1681  
QY 240 MKK 242  
DB 1682 MKR 1684

RESULT 6  
ID CA54 CANFA STANDARD; PRT; 754 AA.  
AC Q28247.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Collagen alpha 5(IV) chain (Fragment).  
GN COL4A5.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]

SEQUENCE FROM N.A.  
RC STRAIN-Sanoyed; TISSUE-Kidney;  
RX MEDLINE=9424866; PubMed=8171024;  
RA Zheng K., Thorne P.S., Marrano P., Baunil R., McInnes R.R.;  
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for  
RT human X-linked hereditary nephritis resulting from a single base  
RT mutation in the gene encoding the alpha 5 chain of collagen type  
RT IV";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire',  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of  
CC canine X-linked hereditary nephritis (HN), a disease similar to  
CC that in humans (also referred to as Alport syndrome) characterized  
CC by progressive renal failure and neurosensory deafness.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U07888; A3560258.1; -.  
PIR; A55267; A55267. C1g helix.  
InterPro; IPR008161; C1g helix.  
InterPro; IPR008160; Collagen.  
InterPro; IPR001442; Procollagen4\_C.  
Pfam; PF01413; C4; 2.  
Pfam; PF01391; Collagen; 8.  
ProDom; PD000007; C1g helix; 1.  
ProDom; PD003923; ProcollagenC4; 1.  
SMART; SM00111; C4; 2.  
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
FT NON TER 1 1  
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.  
FT DOMAIN 531 >754 NON-HELICAL REGION (NC1).  
FT DISULFID 552 643 OR 640 (BY SIMILARITY).  
FT DISULFID 585 640 OR 643 (BY SIMILARITY).  
FT DISULFID 597 603 BY SIMILARITY.  
FT DISULFID 662 ? OR 754 (BY SIMILARITY).  
FT DISULFID 696 754 BY SIMILARITY.  
FT DISULFID 708 714 BY SIMILARITY.  
FT NON TER 754 754  
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 68.8%; Score 922.5; DB 1; Length 754;  
Best Local Similarity 68.4%; Pred. No. 4.1e-78;  
Matches 162; Conservative 32; Mismatches 42; Indels 1; Gaps 1;

QY 1 GLKGRKDGSGSPATWT-TRGFVFTRHSSQTTPAIPSCPGTFLVSGFSLVQGNRAHQ 59  
DB 518 GPDGMQGGPPGPGTSSVAHGFLLTRHSQTTPAIPSCPGTFLVSGFSLVQGNRAHQ 577



```

Db 1635 SRCVCEVPTQITAVHSQDTSVQCPCGCHSGMTGYSFVWHTDAGAGTGQSLOSQSGSCL 1694
Qy 180 EFRASPFLECHGRTGCTNYNSYSFWLASLNERMPKPIPTVTKAGELEKIISRCQVC 239
Db 1695 EFRVAFIECHGRTGCTNYATNHGWFPSIVDQKQFRKPMSTLKGAGLKKRVSRCQVC 1754
Qy 240 MKKR 243
Db 1755 LQNR 1758

RESULT 8
CA24_CASCU STANDARD; PRT; 1763 AA.
ID CA24_CASCU
AC P27393
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS I AND II).
RX MEDLINE=91340768; PubMed=1714907;
RA Pettitt J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
RL J. Biol. Chem. 266:16149-16156 (1991).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I;
CC IsoId=P27393-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P27393-2; Sequence=VSP_001159;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67507; AAA19014.1; -.
DR PIR; S16366; S16366.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR01442; Procollagen4_C.
DR Pfam; PFO1413; C4; 2.
DR Pfam; PFO1391; Collagen; 25.
DR ProDom; PD000607; C1g helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW

```

```

KW Alternative splicing: Glycoprotein; Signal.
FT SIGNAL 1 26 POTENTIAL
FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 27 42 7S DOMAIN.
FT DOMAIN 27 42 TRIPLE-HELICAL REGION.
FT DOMAIN 43 1529 NONHELICAL REGION (NC1).
FT DISULFID 1530 1763 OR 1634 (BY SIMILARITY).
FT DISULFID 1548 1637 OR 1637 (BY SIMILARITY).
FT DISULFID 1581 1634 BY SIMILARITY.
FT DISULFID 1593 1599 BY SIMILARITY.
FT DISULFID 1596 1752 OR 1749 (BY SIMILARITY).
FT DISULFID 1596 1752 OR 1752 (BY SIMILARITY).
FT DISULFID 1690 1749 BY SIMILARITY.
FT DISULFID 1702 1709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 O-LINKED (XYL. . .) (GLYCOSAMINOGLYCAN).
FT CARBOHYD 249 249 (IN ISOFORM II) (POTENTIAL).
FT VARSPLIC 230 266 GEQGRGPGPGPGPVSTGAKGTIGPEGAPGMKGEK ->
FT isoform I).
FT isoform II).
SQ SEQUENCE 1763 AA; 168526 MW; 304F528B8C06AAE0D CRC64;
Query Match 59.4%; Score 796.5; DB 1; Length 1763;
Best Local Similarity 57.9%; Pred. No. 6e-66;
Matches 140; Conservative 41; Mismatches 58; Indels 3; Gaps 2;
Qy 1 GLKXRGDSGSPATWTTTRGTFVTHSQTTPAIPSCPEGTVPVLYSGFSFLVQGNORAGQD 60
Db 1515 GLPQSGPGPPGPGPSYKDGFLLVKHSQTSVEPQCPGGMVWLWDGYSLLYIEGNEKSHNQD 1574
Qy 61 LGTLGSLQRTTTPFLFCNVNDVNCNPNASNDYSYWLSTPALMPMNPAPITGRALEPVIS 120
Db 1575 LQHAGSLSRFTTPFLFCNVNVNVCNPNASNDYSYWLSTPA--PIPMFVSEGGIEPVIS 1632
Qy 121 RCTVCEGPAIAVAHSQTDIPPCPHGWSLWKGFSLMFTSAGSEGTGQALASPGSCLE 180
Db 1693 KCAVCEAPNVAHVSQTIQIPNCPNGWNSLWISYFAMHTGACAGGGGSLSPGSCLE 1692
Qy 181 EFRASPFLECHG-RGTCNYNSYSFWLASLNERMPKPIPTVTKAGELEKIISRCQVC 239
Db 1693 DFRATPFIECNGARGTCHYFANKFSFWLTTIEDQQRIPSEFLKAGSLATRVSRQVC 1752
Qy 240 MK 241
Db 1753 IR 1754
RESULT 9
CA24_CAEEL STANDARD; PRT; 1758 AA.
ID CA24_CAEEL
AC P17140; Q19098; Q19099;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (Lethal protein 2).
GN LET-2 OR CLB-1 OR F01G12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A., AND FUNCTION.
RP STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene."
RL J. Cell Biol. 123:255-264 (1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RX STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen

```



RA Pihlajaniemi T., Kurkinen M.;  
RT "The complete primary structure of mouse alpha 2(IV) collagen."  
RL Alignment with mouse alpha 1(IV) collagen."  
RN J. Biol. Chem. 264:6318-6324(1989).  
[2]  
RP SEQUENCE OF 1-33 FROM N.A.  
RX MEDLINE=89066738; PubMed=3196626;  
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
RT "Head-to-head arrangement of murine type IV collagen genes."  
RN J. Biol. Chem. 263:19274-19277(1988).  
[3]  
RP SEQUENCE OF 970-1480 FROM N.A.  
RX MEDLINE=86220192; PubMed=301132;  
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,  
RT Deutzmann R., Timpl R., Kuehn K.;  
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain."  
RL Eur. J. Biochem. 157:49-56(1986).  
[4]  
RP SEQUENCE OF 1480-1707 FROM N.A.  
RX MEDLINE=87054581; PubMed=3780963;  
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;  
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV)."  
RN FEBS Lett. 208:203-207(1986).  
[5]  
RP SEQUENCE OF 1481-1707 FROM N.A.  
RX MEDLINE=87250460; PubMed=3597383;  
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
RT Saus J., Pihlajaniemi T.;  
RT "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen."  
RL J. Biol. Chem. 262:8496-8499(1987).  
[6]  
RP SEQUENCE OF 1041-1489 FROM N.A.  
RX MEDLINE=87005245; PubMed=3758345;  
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;  
RT "Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen."  
RN FEBS Lett. 206:29-32(1986).  
[7]  
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.  
RX MEDLINE=85296379; PubMed=3839908;  
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;  
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene."  
RN Nature 317:177-179(1985).  
[8]  
RP SEQUENCE OF 1-60 FROM N.A.  
RX MEDLINE=89071759; PubMed=3200851;  
RA Burdello P.D., Martin G.R., Yamada Y.;  
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer."  
RN Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
CC -----  
DR EMBL; M23334; AAA51626.1; JOINED.  
DR EMBL; M23333; AAA51626.1; JOINED.  
DR EMBL; J04695; AAA50293.1; -  
DR EMBL; J04448; AAA37438.1; -  
DR EMBL; X04647; CAA28308.1; -  
DR EMBL; M15833; AAA37341.1; -  
DR EMBL; X04410; CAA27998.1; -  
DR EMBL; X02896; CAA26655.1; -  
DR EMBL; X02897; CAB51614.1; -  
DR EMBL; X02898; CAA26657.1; -  
DR EMBL; X02899; CAA26658.1; -  
DR PIR; A33526; A33526.  
DR MGI; MGI-88458; Col4a2.  
DR CO; CO10005604; C:basement membrane; IDA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR ProDom; PD000007; C1g\_helix; 7.  
DR ProDom; PD003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Signal.  
FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.  
FT CHAIN 184 1707 TRIPLE-HELICAL REGION.  
FT DOMAIN 184 1479 NONHELICAL REGION (NC1).  
FT DISULFID 1490 1707 OR 1585 (BY SIMILARITY).  
FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).  
FT DISULFID 1544 1550 BY SIMILARITY.  
FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).  
FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).  
FT DISULFID 1653 1660 BY SIMILARITY.  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 1051 1051 P -> R (IN REF. 6).  
FT CONFLICT 1097 1097 S -> G (IN REF. 7).  
FT CONFLICT 1171 1171 G -> S (IN REF. 6).  
FT CONFLICT 1179 1179 P -> R (IN REF. 6).  
FT CONFLICT 1241 1241 Q -> E (IN REF. 6).  
FT CONFLICT 1328 1328 P -> L (IN REF. 4).  
FT CONFLICT 1573 1573 V -> L (IN REF. 4).  
FT CONFLICT 1623 1623 Y -> H (IN REF. 4).  
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;  
Query Match 56.8%; Score 760.5; DB 1; Length 1707;  
Best Local Similarity 57.6%; Pred. No. 1.3e-62;  
Matches 140; Conservative 37; Mismatches 61; Indels 5; Gaps 4;  
QY 1 GLKRGKSGSPATWTRGFVTRHSQTALPSCPCTGVPLYSFSLFVQGNRAHQD 60  
1466 GRGSGPLGMPGRSVISGYLLVSKHSQTDQEPMPGVGNKLVSGYLLYFEGQEKHQD 1525  
DB 61 LGTLGSLQRTMTPLFCNVNDVNCNFRNDYSYLSLTPALPMNMAPIGTRALEPVIS 120  
1526 LGLAGSLARFTNPLFLYCNPDGVYASRNDKSYLSLTTA--PLPMNPAVEEIKPVIS 1583  
QY 121 RCTVCGPAIAVAHSQTDDIPCPHGMIKSGSFIMFTSAGSEGTQALASPGSCLE 180  
1584 RSVCEAPAVAVHSQDTSIHCPCAGWRSLSWISFLMYTAGDEGGQSIVSPGSCLE 1643  
QY 181 EFRASPFLECH-GRGTCNTYNSYSFSLASLNPENFRK-PIPSTVKGAELEKIISRCOV 238

```
Db 1644 DFRATPIECNGRGTCYFANKYSFWLTTI-PEQNQFQSPSADTLKAGLIRTHSRCQV 1702
QY 239 CMK 241
Db 1703 CMK 1705

RESULT 11
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
RT collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19408-19493(1988).
RN [2]
RP SEQUENCE OF 1-1042 FROM N.A.
RX MEDLINE=89066769; PubMed=3345760;
RA Hostikka S.L., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence
RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
RT chain reveals deletions in the alpha 1(IV) chain.";
RL Eur. J. Biochem. 172:35-42(1988).
RN [3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
RT chain CDNA reveals extensive homology with the NC-1 domain of alpha 1
RT (IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286(1987).
RN [4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-515(1987).
RN [5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [7]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
RT basement membrane collagen type IV are arranged head-to-head and
```

```
RT separated by a bidirectional promoter of unique structure.";
RL EMBO J. 7:2687-2695(1988).
RN [8]
RP SEQUENCE OF 1-33 FROM N.A.
RX TISSUE=Skin;
RC MEDLINE=93305049; PubMed=8317999;
RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter
RT region of human collagen type IV genes, involved in the regulation of
RT divergent transcription.";
RL Biochem. J. 292:687-695(1993).
RN [9]
RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RX TISSUE=Placenta;
RC MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
RN [10]
RP FUNCTION: Type IV collagen is the major structural component of
RP glomerular basement membranes (GBM), forming a 'chicken-wire'
RP meshwork together with laminins, proteoglycans and entactin/
RP nidogen.
RN [11]
RP SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
RP alpha 6(IV), each of which can form a triple helix structure
RP with 2 other chains to generate type IV collagen network.
RN [12]
RP DOMAIN: Alpha chains of type IV collagen have a noncollagenous
RP domain (NC1) at their C-terminus, frequent interruptions of the
RP G-X-Y repeats in the long central triple-helical domain (which may
RP cause flexibility in the triple helix), and a short N-terminal
RP triple-helical 7S domain.
RN [13]
RP PTM: Prolines at the third position of the tripeptide repeating
RP unit (G-X-Y) are hydroxylated in some or all of the chains.
RN [14]
RP PTM: Type IV collagens contain numerous cysteine residues which
RP are involved in inter- and intramolecular disulfide bonding. 12 of
RP these, located in the NC1 domain, are conserved in all known type
RP IV collagens.
RN [15]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RN [16]
RP EMBL; X05562; CAA29076.1; -
RP EMBL; X05610; CAA29098.1; -
RP EMBL; J02780; AAA58422.1; -
RP EMBL; M36963; AAA53099.1; -
RP EMBL; X12784; CAA31275.1; -
RP EMBL; J04217; AAA53097.1; -
RP PIR; A32024; CGHU2B.
RP Genew; HGNC:2203; COL4A2.
RN [17]
RP MIM; 120090; -
RP GO; GO:0005587; C:collagen type IV; TAS.
RP GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
RP GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
RN [18]
RP InterPro; IPR008161; Clg helix.
RP InterPro; IPR008160; Collagen.
RP InterPro; IPR001442; Procollagen4_C.
RP Pfam; PF01413; C4; 2.
RP Pfam; PF01391; Collagen; 24.
RP ProDom; PD000007; Clg helix; 7.
RP ProDom; PD003923; ProcollagenC4; 1.
RP SMART; SM00111; C4; 2.
RN [19]
RP Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
RP Glycoprotein; Basement membrane; Collagen; Signal.
RN [20]
RP SIGNAL 1 25
RP PROPEP 26 183
FT CHAIN 184 1712
FT DOMAIN 184 1484
RN [21]
RP AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
RN [22]
RP COLLAGEN ALPHA 2(IV) CHAIN.
RN [23]
RP TRIPLE-HELICAL REGION.
```



```
FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1537 1555 OR 1593 (BY SIMILARITY).
FT DISULFID 1549 1555 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
FT CONFLICT 471 471 R -> P (IN REF. 2).
FT CONFLICT 683 683 A -> G (IN REF. 2).
FT CONFLICT 1575 1575 M -> I (IN REF. 5).
FT CONFLICT 1663 1663 G -> H (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 56.4%; Score 755.5; DB 1; Length 1712;
Best Local Similarity 58.08; Pred. No. 3.8e-62;
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;

Qy 1 GLKGRGDSGSPATWTRGRFVTRHSQTTPAIPCEGTVPYSGFSLFVQGNQRAHQD 60
Db 1471 GRPGSPGLPGMGSRVSGVLLVKSQTDQBPMCPVGMNKLWSGYSLLYFEGQEKAHNQD 1530
Qy 61 LGTIGSCLORTTTPFLFCNVNDVNCNPNASNDYSYWLSTALMPNMAPITGRALSPYIS 120
Db 1531 LGLAGSCLARFSTPFYLCNPDGVCYASRNDKSYWLSTTA--PLPMMPVAEDEIKPYIS 1588
Qy 121 RCTVCEGPALAIHVSTQTDIPPCPHGWI SLWKGFSPIMFTSAGSEGTQALASPGSCLE 180
Db 1589 RCVCEAPALAIHVSDVSIHPCHPACGRSLWIGYSFLMHTAAGDEGGQSLVSPGSCLE 1648
Qy 181 EFRASPLECH-GRGTQNYNSNSVFWLASINLPRMPR-KPISTVKGAELEKILRCQV 238
Db 1649 DFRATPFIECNGRGRTCHYANKYSFWLTTI-PEQSPQGSFSDTLKAGLIRTHIRGQV 1707
Qy 239 CMK 241
Db 1708 CMK 1710

RESULT 12
CA64 HUMAN STANDARD; PROT: 1691 AA.
ID CA64 HUMAN STANDARD; PROT: 1691 AA.
AC Q14031; Q12823; Q14053; Q9NQX5; Q9NTX3; Q9UJ76; Q9UMG6; Q9V4L4;
DI 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eye, and Kidney;
RX MEDLINE=94171779; PubMed=8125972;
RA Ohtashi T., Sugimoto M., Matsu M.-G., Ninomiya Y.;
RT Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.;
RL J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Readers S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RL J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
```

```
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Readers S.T., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT in Alport syndrome-associated leiomyomatosis.";
RN Genomics 33:473-479(1996).
RL [4]
RP SEQUENCE FROM N.A.
RA Bird C., Grahame D., Lawlor S., Wilson S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).
RX MEDLINE=93361972; PubMed=8356449;
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,
RA de Paep A., Tryggvason K., Readers S.T.;
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in
RT inherited smooth muscle tumors.";
RL Science 261:1167-1169(1993).
RL Science 261:1167-1169(1993).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=Q14031-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q14031-2; Sequence=VSP_001174;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; D21337; BAA04809.1; -
CC EMBL; U04845; AAA19569.2; -
CC EMBL; U47004; AAB19038.1; -
CC EMBL; U46959; AAB19038.1; JOINED.
CC EMBL; U46961; AAB19038.1; JOINED.
CC EMBL; U46962; AAB19038.1; JOINED.
CC EMBL; U46963; AAB19038.1; JOINED.
CC EMBL; U46964; AAB19038.1; JOINED.
CC EMBL; U46965; AAB19038.1; JOINED.
CC EMBL; U46966; AAB19038.1; JOINED.
CC EMBL; U46967; AAB19038.1; JOINED.
CC EMBL; U46968; AAB19038.1; JOINED.
CC EMBL; U46969; AAB19038.1; JOINED.
CC EMBL; U46970; AAB19038.1; JOINED.
CC EMBL; U46971; AAB19038.1; JOINED.
CC EMBL; U46972; AAB19038.1; JOINED.
CC EMBL; U46973; AAB19038.1; JOINED.
CC EMBL; U46974; AAB19038.1; JOINED.
CC EMBL; U46975; AAB19038.1; JOINED.
CC EMBL; U46976; AAB19038.1; JOINED.
CC EMBL; U46977; AAB19038.1; JOINED.
```

```

DR ENBL; AL136080; CAB96748.1; -.
DR ENBL; AL031177; CAA20120.1; -.
DR ENBL; L22763; AAA16338.1; -.
DR PIR; A54122; CGHU68.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0003201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; F:extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn_4.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR KW Extracellular matrix; Connective tissue; Basement membrane;
DR KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 55.4%; Score 739; DB 1; Length 1691;
Best Local Similarity 54.7%; Pred. No. 1.3e-60;
Matches 133; Conservative 42; Mismatches 64; Indels 4; Gaps 3;

Qy 1 GLKGRDSSPATWTRGFTVFRHSOTTAIPSCPEGTVLYSGFSLFVQGNQRAHQD 60
Db 1449 GQQQGFPMQPMQPMQPMQPMQPMQPMQPMQPMQPMQPMQPMQPMQPMQPM 1508
Qy 61 LGTLGSLCQRFTWMPFLFCNVNDYCNFASNDYSYWLSTPALMPMNPVITGRALEPYIS 120
Db 1509 LGPAGSCLPFTSWTFYICNINEVCHYARNDKSYWLSTTA--PIPMVPYSQTPIQV 1566
Qy 121 RCTVCEGPAITAIVHSQTTDIPPCPHQWISLWKGFSFIMFTSAGSEGTQALASPGSCLE 180
Db 1567 RCSVCEAPSOAIAVHSQDITIPQCLGWRSLWIGYSFLMHTAAGGGGOSLVSPGSCLE 1626
Qy 181 EFRASPFLCHG-RGTCNYSNSYSFWMASLNNPFRK-PIPTVKAGELEKIIISRCQV 238
Db 1627 DFRATPIECSGARGTCHYFANKYSFWLTVBERQQGELPVSETLKAGQLHTVRSQV 1686
Qy 239 CMK 241
Db 1687 CMK 1689

RESULT 13
CA44_BOVIN STANDARD; PRT; 453 AA.
ID CA44_BOVIN
AC Q29442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=921112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Redders S.T.;
RT "the alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens.";
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]

```





DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 5.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM01111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1  
 FT DOMAIN <1 392 TRIPLE-HELICAL REGION.  
 FT DOMAIN 393 623 NON-HELICAL REGION (NC1).  
 FT DISULFID 413 502 OR 499 (BY SIMILARITY).  
 FT DISULFID 446 499 OR 502 (BY SIMILARITY).  
 FT DISULFID 458 464 BY SIMILARITY.  
 FT DISULFID 521 619 OR 616 (BY SIMILARITY).  
 FT DISULFID 555 616 OR 619 (BY SIMILARITY).  
 FT DISULFID 567 574 BY SIMILARITY.  
 SQ SEQUENCE 623 AA; 62393 MW; CBC9BB31242FE92 CRC64;  
 Query Match 52.4%; Score 702; DB 1; Length 623;  
 Best Local Similarity 49.0%; Pred. No. 1.1e-57;  
 Matches 128; Conservative 45; Mismatches 56; Indels 22; Gaps 5;  
 Qy 1 GLKRGDSGSPAF-----WTRGFVFRHSQTTAIPSCPGTVPVLYS 43  
 Db 363 GHKGDTCGAGRPGAPGPPGTGDPKXGLGPGVLSGFLVLSQTDQEPACPMGMPRLWT 422  
 Qy 44 GFSFLFVQGNORAGQDLGTLSCLQRTFTMPFLFCNVNVDVCFNFSRNDYSYWLSTPALM 103  
 Db 423 GYSLLYEGEKEKANKQDLGAGSLPFTLPAYCNHQVCHYQNRNDKSYWLASAG-- 480  
 Qy 104 PMNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPCPGWSLWKGFSFIMFTSA 163  
 Db 481 PLPMPLSEBIEIRYISKVACEAPAAQVAVHSQDSQIPCPRAWRLWIGYSFLMHTGA 540  
 Qy 164 GSEGTQALASPGSCLEEFASPLKCHGR-GTCNYSNSYSFWLASLNPDE-RMFRKPDP 221  
 Db 541 GDQGGQALWSPGSCLEDFRAAPLECGRQGTCHFFANYSFWLTPVDPDLQVFSAPS 600  
 Qy 222 STVRAGLEK-IISRCQVCMK 241  
 Db 601 DTLKESQAQRKISRCQVCVK 621  
 RESULT 15  
 CA44 HUMAN  
 ID CA44 HUMAN STANDARD; PRT; 1690 AA.  
 AC P53420.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Collagen alpha 4(IV) chain precursor.  
 GN COL4A4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=95014445; PubMed=7523402;  
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reenders S.T.;  
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)  
 RT chain. Comparison with structure and expression of the other alpha  
 RT (IV) chains.";  
 RT J. Biol. Chem. 269:26172-26177(1994).  
 RL [2]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Moncton R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome

RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RP [3]  
 RP SEQUENCE OF 1219-1690 FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=93374047; PubMed=8365481;  
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;  
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)  
 RT collagen chain.";  
 RL FEBS Lett. 330:122-128(1993).  
 RP [4]  
 RP SEQUENCE OF 1407-1507 FROM N.A.  
 RX MEDLINE=93054733; PubMed=1429714;  
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 RT alpha 4 chain of basement membrane collagen type IV and assignment of  
 RT the gene to the distal long arm of human chromosome 2.";  
 RL J. Biol. Chem. 267:23753-23758(1992).  
 RP [5]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RP [6]  
 RP VARIANT AS SER-1201.  
 RX MEDLINE=95078927; PubMed=7987396;  
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
 RA Smeets H.J.M., Reenders S.T.;  
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)  
 RT collagen genes in autosomal recessive Alport syndrome.";  
 RL Nat. Genet. 8:77-82(1994).  
 RP [7]  
 RP VARIANT FBH GUJ-897.  
 RX MEDLINE=96379660; PubMed=8787673;  
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
 RT "Benign familial hematuria due to mutation of the type IV collagen  
 RT alpha4 gene.";  
 RL J. Clin. Invest. 98:1114-1118(1996).  
 RP [8]  
 RP VARIANTS AS, AND VARIANTS.  
 RX MEDLINE=99011253; PubMed=9792860;  
 RA Boye E., Wollet G., Forestier L., Cohen-Solal L., Heidet L.,  
 RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;  
 RT "Determination of the genomic structure of the COL4A4 gene and of  
 RT novel mutations causing autosomal recessive Alport syndrome.";  
 RL Am. J. Hum. Genet. 63:1329-1340(1998).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
 CC alpha 6(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- SUBCELLULAR LOCATION: Cell surface (potential).  
 CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
 CC colocalized and present only in basement membranes of kidney, eye,  
 CC cochlea, lung and brain.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -!- DISASPS: Defects in COL4A4 are a cause of autosomal recessive  
 CC Alport syndrome (AS) [MIM:203780], an hereditary disorder



GenCore version 5.1.6  
Copyright (C) 1993 - 2004: CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 146.518 Seconds  
(without alignment)  
525.440 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKGRGSGSPATWTRGF.....KAGELEKIISRQVQVCKKRH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	245	4 Q9NYC4	Q9NYC4 homo sapien
2	1221.5	91.2	246	11 Q61435	Q61435 mus musculus
3	1221.5	91.2	1669	11 Q9QZ50	Q9QZ50 mus musculus
4	1164	86.9	230	11 Q63122	Q63122 rattus norv
5	1158	86.4	212	6 Q28512	Q28512 macaca mula
6	1103	82.3	212	6 Q28567	Q28567 ovis aries
7	1084	80.9	210	6 Q28273	Q28273 canis fami
8	1072	80.0	203	6 Q28682	Q28682 oryctolagus
9	1055	78.7	203	6 Q29032	Q29032 sus scrofa
10	960	71.6	1075	4 Q86X41	Q86X41 homo sapien
11	960	71.6	1621	4 Q9H4R9	Q9H4R9 homo sapien
12	957.5	71.5	979	13 Q919K3	Q919K3 gallus gall
13	952.5	71.1	253	11 Q61436	Q61436 mus musculus
14	952.5	71.1	585	11 Q80V57	Q80V57 mus musculus
15	952.5	71.1	799	11 Q8BNS7	Q8BNS7 mus musculus
16	952.5	71.1	1691	11 Q9ESQ2	Q9ESQ2 mus musculus

17	943.5	70.4	886	4 Q9NIB7	Q9NIB7 homo sapien
18	939	70.1	229	4 Q9NYC5	Q9NYC5 homo sapien
19	930	69.4	226	11 Q9BLQ8	Q9BLQ8 mus musculus
20	928	69.3	229	4 Q8NFB8	Q8NFB8 homo sapien
21	924.5	69.0	1684	6 Q8HYC1	Q8HYC1 canis fami
22	906.5	67.6	1688	6 Q86622	Q86622 canis fami
23	893	66.6	225	6 Q28271	Q28271 canis fami
24	836	62.4	1752	5 Q07265	Q07265 strongyloce
25	834	62.2	161	11 Q61430	Q61430 mus musculu
26	776	57.9	1802	5 Q17153	Q17153 brugia mala
27	773	57.7	179	11 P70165	P70165 mus musculu
28	767.5	57.3	1747	5 Q26640	Q26640 strongyloce
29	757.5	56.5	673	4 Q14052	Q14052 homo sapien
30	756.5	56.5	358	11 Q91VI3	Q91VI3 mus musculu
31	735	54.9	1691	11 Q9ESQ1	Q9ESQ1 mus musculu
32	734	54.8	545	11 Q9XK97	Q9XK97 mus musculu
33	710	53.0	1775	5 Q9VMV4	Q9VMV4 drosophila
34	706	52.7	1682	11 Q9QZR9	Q9QZR9 mus musculu
35	705	52.6	312	11 Q64457	Q64457 mus musculu
36	693.5	51.8	1024	5 Q8T7S4	Q8T7S4 anopheles g
37	683.5	51.0	1723	5 Q9GQB1	Q9GQB1 hydra atten
38	680.5	50.8	202	6 Q28272	Q28272 canis fami
39	670	50.0	205	6 Q28274	Q28274 canis fami
40	649	48.4	208	6 Q29468	Q29468 canis fami
41	648.5	48.4	713	5 Q9GV24	Q9GV24 sarcophaga
42	646.5	48.2	1761	5 Q18407	Q18407 drosophila
43	646.5	48.2	1940	5 Q9VMV5	Q9VMV5 drosophila
44	596	44.5	854	5 Q09238	Q09238 pseudocorti
45	585	43.7	112	11 Q8CCD6	Q8CCD6 mus musculu

#### ALIGNMENTS

#### RESULT 1

Q9NYC4	PRELIMINARY;	PRT;	245 AA.
ID Q9NYC4			
AC Q9NYC4;			
DT 01-OCT-2000 (Tremblrel. 15, Created)			
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE Tumstatin (Fragment).			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,			
RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R., derived			
RT 'Distinct anti-tumor properties of a type IV collagen domain derived			
RL from basement membrane.';			
RL J. Biol. Chem. 0:0-0(2000).			
DR EMBL; AF258351; AAF72632.1; ..			
DR GO; GO:0005581; C:collagen; IEA.			
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.			
DR GO; GO:0003676; P:nucleic acid binding; IEA.			
DR InterPro; IPR001442; Procollag4_C.			
DR InterPro; IPR000504; RNA_rec_mot.			
DR Pfam; PF01413; C4; 2.			
DR ProDom; PD003923; ProcollagNC4; 1.			
DR SMART; SM00111; C4; 2.			
DR PROSITE; PS00030; RRM_RNP_1; 1.			
FT NON_TER			
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;			

Query Match 100.0%; Score 1340; DB 4; Length 245;  
Best Local Similarity 100.0%; Pred. No. 2.3e-129;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGSGSPATWTRGFVTRHSQTATPSCPECTVPLYSFSLFVQGNQRAQCQD 60

DB 2 GLKGRGSGSPATWTRGFVTRHSQTATPSCPECTVPLYSFSLFVQGNQRAQCQD 61

```

QY 61 LGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYIS 120
DB 62 LGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYIS 121
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCLE 180
DB 122 RCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSCLE 181
QY 181 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 240
DB 182 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 241
QY 241 MKKH 244
DB 242 MKKH 245

RESULT 2
ID Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: Sequence, distribution, association with laminins, and
RT developmental switches.";
RL J. Cell Biol. 127:879-891(1994).
EN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CA84529.1; -.
DR PIR; I48302.
DR MGI; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
FT NON_TER
SQ SEQUENCE 246 AA; 26993 MW; A9E5434F5836F324 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 246;
Best Local Similarity 90.6%; Pred. No. 3.4e-117;
Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKGRGSGSPATWT-TRGFVFRHSQTTAIPSCPGTVPVLYSGFSFLFVQGNRAHQ 59
DB 2 GLKGNPGRGTPTATGTRVRGFIIFRHSQTTAIPSCPGTQPLYSGFSLLFVQGNKRAHQ 61
QY 60 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 119
DB 62 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 121
QY 120 SRTCTCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSC 179
DB 122 SRTCTCEGPAIAIAVHSQTTAIPCPQDWSLWKGRFSIMFTSAGSEGTGOALASPGSC 181
QY 180 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 239
DB 182 SRTCTCEGPAIAIAVHSQTTAIPCPQDWSLWKGRFSIMFTSAGSEGTGOALASPGSC 181

Query Match 91.2%; Score 1221.5; DB 11; Length 246;
Best Local Similarity 90.6%; Pred. No. 3.4e-117;
Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKGRGSGSPATWT-TRGFVFRHSQTTAIPSCPGTVPVLYSGFSFLFVQGNRAHQ 59
DB 1425 GLKGNPGRGTPTATGTRVRGFIIFRHSQTTAIPSCPGTQPLYSGFSLLFVQGNKRAHQ 1484
QY 60 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 119
DB 1485 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 1544
QY 120 SRTCTCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSC 179
DB 1545 SRTCTCEGPAIAIAVHSQTTAIPCPQDWSLWKGRFSIMFTSAGSEGTGOALASPGSC 1604
QY 180 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 239
DB 1605 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 1664
QY 240 MKKH 244
DB 1665 MKKH 1669

```

```

QY 180 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 239
DB 182 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 241
QY 240 MKKH 244
DB 242 MKKH 246

RESULT 3
ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.
AC Q9QZS0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha 3 collagen IV.
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169387; AAD50449.1; -.
DR PIR; I48302; I48302.
DR MGI; MGI:104688; Col4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RNP_RNP_1; 1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 1669;
Best Local Similarity 90.6%; Pred. No. 3.3e-116;
Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKGRGSGSPATWT-TRGFVFRHSQTTAIPSCPGTVPVLYSGFSFLFVQGNRAHQ 59
DB 1425 GLKGNPGRGTPTATGTRVRGFIIFRHSQTTAIPSCPGTQPLYSGFSLLFVQGNKRAHQ 1484
QY 60 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 119
DB 1485 DLGTGLSCLOQFTTTPFPCNVNDVNCNPNASNDYSYWLSTPALMPMNAIPITGRALEPYI 1544
QY 120 SRTCTCEGPAIAIAVHSQTTDIPCPHGWSLWKGRFSIMFTSAGSEGTGOALASPGSC 179
DB 1545 SRTCTCEGPAIAIAVHSQTTAIPCPQDWSLWKGRFSIMFTSAGSEGTGOALASPGSC 1604
QY 180 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 239
DB 1605 EFRASPFLCHGRGTCNYSNSYFWLASLNPERMFRKPIPSYKAGLEKIISRCQVCM 1664
QY 240 MKKH 244
DB 1665 MKKH 1669

```

```
RESULT 4
Q63122 ID Q63122 PRELIMINARY; PRT; 230 AA.
AC Q63122;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RX MEDLINE=98210005; PubMed=9550634;
RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;
RT "Sequence analysis of the 'Goodpasture antigen' of mammals.";
RL Nephrol. Dial. Transplant. 13:602-607(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47281; AAB72238.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 230
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 86.9%; Score 1164; DB 11; Length 230;
Best Local Similarity 91.6%; Pred. No. 2.6e-111;
Matches 208; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

Qy 18 RGFVTRHSQTTPSPCEGTPLYSQFSLFVQGNQRAHGQDLGLTGLSCLQRTFTMPPL 77
Db 4 RGFVTRHSQTTPSPCEGTPLYSQFSLFVQGNQRAHGQDLGLTGLSCLQRTFTMPPL 63

Qy 78 FCNVNDVNCNPNRNDYSLWSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQ 137
Db 64 FCNVNDVNCNPNRNDYSLWSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQ 123

Qy 138 TTPIPPCHGWISLWKGFSFIMFTSAGSEGTQALASPGSCLEEFPRASPFLECHGRGTGN 197
Db 124 TTAIPPCQGWISLWKGFSFIMFTSAGSEGTQALASPGSCLEEFPRASPFLECHGRGTGN 183

Qy 198 YVNSYSFWLASLNPERMFKPSTVTKAGELEKIIISRCQVCMKKRH 244
Db 184 YVNSYSFWLASLNPERMFKPSTVTKAGELEKIIISRCQVCMKKRH 230

RESULT 5
Q28512 ID Q28512 PRELIMINARY; PRT; 212 AA.
AC Q28512;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.

Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47280; AAA91861.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
FT NON_TER 212
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;

Query Match 86.4%; Score 1158; DB 6; Length 212;
Best Local Similarity 99.1%; Pred. No. 9.6e-111;
Matches 210; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 33 SCPEGTVPLXSGFSLFVQGNQRAHGQDLGLTGLSCLQRTFTMPPLFCNVNDVNCNFSRND 92
Db 1 SCPEGTVPLXSGFSLFVQGNQRAHGQDLGLTGLSCLQRTFTMPPLFCNVNDVNCNFSRND 60

Qy 93 YSYWLSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLW 152
Db 61 YSYWLSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLW 120

Qy 153 KGRSFIMFTSAGSEGTQALASPGSCLEEFPRASPFLECHGRGTCTNYNSYSFWLASLNP 212
Db 121 KGRSFIMFTSAGSEGTQALASPGSCLEEFPRASPFLECHGRGTCTNYNSYSFWLASLNP 180

Qy 213 ERMRKPIPTVTKAGELEKIIISRCQVCMKKRH 244
Db 191 ERMRKPIPTVTKAGELEKIIISRCQVCMKKRH 212

RESULT 6
Q28567 ID Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

```
DR InterPro; IPR001442; Procollagn4 C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 82.3%; Score 1103; DB 6; Length 212;
Best Local Similarity 92.4%; Pred. No. 4.3e-105;
Matches 195; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVNDVCFASRND 60

QY 93 YSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 152
DB 1 YSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 120

QY 153 KGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 212
DB 121 KGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 180

QY 213 ERMFRKPIPTVKAGELEKIISRCQVCMKKR 243
DB 181 QRMFRKPIPTVKAGELEKIISRCQVCMKKR 211

RESULT 7
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A3 gene mutation."
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4 C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA82633D CRC64;

Query Match 80.9%; Score 1084; DB 6; Length 210;
Best Local Similarity 91.9%; Pred. No. 3.8e-103;
Matches 193; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 23 TRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVN 82
DB 1 TRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVN 60

QY 83 DVCNFAERNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIP 142
DB 1 DVCNFAERNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIP 120

QY 143 PCPHGWISLWKGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNS 202
DB 121 SCPNGWISLWKGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNS 180

QY 203 YSWLASLNPFRMPKPIPTVKAGELEKI 232
DB 181 YSWLASLNPFRMPKPIPTVKAGELEKI 210

RESULT 8
Q28682 PRELIMINARY; PRT; 203 AA.
AC Q28682;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47283; AAA91893.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4 C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 80.0%; Score 1072; DB 6; Length 203;
Best Local Similarity 95.1%; Pred. No. 6.2e-102;
Matches 193; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNQRAHGQDLGTLGSCLCQRTTTPFLFCNVNDVCFASRND 60

QY 93 YSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 152
DB 1 YSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLW 120

QY 153 KGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 212
DB 121 KGFSEFMFTSAGSEGTGOALASPGSCLEEFRAFPFLECHGRGTCNYNSYFWLASLNP 180

QY 213 ERMFRKPIPTVKAGELEKIISRCQVCMKKR 235
DB 181 QRMFRKPIPTVKAGELEKIISRCQVCMKKR 203
```



```
RESULT 9
Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Collagen.
KW
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 78.7%; Score 1055; DB 6; Length 203;
Best Local Similarity 93.1%; Pred. No. 3.5e-100;
Matches 189; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 33 SCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLGSLQRTTTPFLFCNVNDVCFASRND 92
DB 1 SCPEGTVPVLYSGFSLFVQGNRSHGQDGLTGLGSLQRTTTPFLFCNVNDVCFASRND 60
QY 93 YSYWLSTPALMPMNAITGRALBPYISRCTVCEGPAIAIAVHSQTTDPPCPHGWISLW 152
DB 61 YSYWLSTPALMPMNAITGRALBPYISRCTVCEGPAIAIAVHSQTTDPPCPHGWISLW 120
QY 153 KGFSFIMFTSAGSGTGQALASPGSCLEEFRAFPFLECHGRGTCNYYNSYSFWLASLNP 212
DB 121 KGFSFIMFTSAGSGTGQALASPGSCLEKFRASPFIECHGRGTCNYYNSYSFWLASLND 180
QY 213 ERMRKPIPTVXAGELEKIISR 235
DB 181 ERMRKPIPTVXAGELEKIISR 203

RESULT 10
Q86X41 PRELIMINARY; PRT; 1075 AA.
AC Q86X41;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
DR Collagen.
KW
FT NON_TER 1
FT NON_TER 1075
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 71.6%; Score 960; DB 4; Length 1075;
Best Local Similarity 69.0%; Pred. No. 1.4e-89;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQD 60
DB 835 GLPGSMGPPGTPS--VDHGFLVTRHSQTIDDPQCPSGKILYHGYSLLYVQGNRAHGQD 892
QY 61 LGTGLSCLQRTTTPFLFCNVNDVCFASRNDYISWLTSPALMPMNAITGRALBPYIS 120
DB 893 LGTAGSCLRKFTSTMPFLFCNVNNVCFASRNDYISWLTSPALMPMNAITGRALBPYIS 952
QY 121 RCTVCEGPAIAIAVHSQTTDPPCPHGWISLWKFIMFTSAGSGTGQALASPGSCLE 180
DB 953 RCACEAPAMVAVHSQTIQIPCPGWSLSLWICYFSFVMTSAGSGTGQALASPGSCLE 1012
QY 181 EFRASPFLECHGRGTCNYYNSYSFWLASLNPFRKPIPTVXAGELEKIISRQVCM 240
DB 1013 EFRSAPFIECHGRGTCNYYANAYFWLATIERSEMFKKPTSTLKAGELRTHVSRQVCM 1072

QY 241 KK 242
DB 1073 RR 1074

RESULT 11
Q9H4R9 PRELIMINARY; PRT; 1621 AA.
ID Q9H4R9
AC Q9H4R9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1 (fragment)).
GN COL4A1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL390755; CAC3153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; C1g_helix; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR Collagen.
KW
FT NON_TER 1
FT NON_TER 1621
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDOEDBA2 CRC64;
```



```

Query Match      71.6%; Score 960; DB 4; Length 1621;
Best Local Similarity 59.0%; Pred. No. 2.3e-89;
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKKGKDGSGPATWTRGFVTRHSQTTAIPSCPEGTVLYSGFSFLVQGNQRAHQD 60
DB 1381 GLPGSMGPPGTPS--VDHGLVTRHSQTTDDPCPSGTRKILYHGYSLLVQGNRAHQD 1438

QY 61 LGTLGSCLOQFTTNPFLFCNVNDVQNFASRNDYSYWLSTPALMPNMMAPIITGRALEPYIS 120
DB 1439 LGTAGSCLRFSTMPFLFCNVNDVQNFASRNDYSYWLSTPALMPNMMAPIITGENTRPFIS 1498

QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFIFIMFTSAGSEGTGOALASPGSCLE 180
DB 1499 RCAVCEAPAMVAHSSQTTQIIPCPSGWSSLMIGYSFVNMHTSAGAEGSQALASPGSCLE 1558

QY 181 EFRASPFLECHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKIIISRCQVCM 240
DB 1559 EFRASPFIECHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKIIISRCQVCM 1618

QY 241 KK 242
DB 1619 RR 1620

RESULT 12
Q919K3 PRELIMINARY; PRT; 979 AA.
ID Q919K3
AC Q919K3
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
basal lamina."
RL Dev. Biol. 0:0-0(2000).
DR EXEL; AF239838; AA244681.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 12.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER
SQ SEQUENCE 1 1
1 5020 MW; 581017D911ED4299 CRC64;

Query Match      71.5%; Score 957.5; DB 13; Length 979;
Best Local Similarity 67.3%; Pred. No. 2.3e-89;
Matches 167; Conservative 33; Mismatches 41; Indels 7; Gaps 1;

QY 1 GLKKGKDGSGSP-----ATWTRGFVTRHSQTTAIPSCPEGTVLYSGFSFLVQGN 53
DB 730 GIPGPPGEGLPFGAMGPPGASVAHGLVTRHSQTTPEPCPGTGLIHYGYSLLVQGN 789

QY 54 QRAHQDLGLGSCLOQFTTNPFLFCNVNDVQNFASRNDYSYWLSTPALMPNMMAPIITGR 113
DB 790 ERAHQDLGTAGSCLRFSTMPFLFCNVNDVQNFASRNDYSYWLSTPEPMPMSMAPIITGB 849

QY 114 ALEPYSRCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFIFIMFTSAGSEGTGOALA 173
DB 114 ALEPYSRCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFIFIMFTSAGSEGTGOALA 173

```

```

Db 850 SIRPFISRCVCEAPAMVIAVHSQTTQIIPCPGWSLWIGYSFVNMHTSAGAEGSQALA 909
QY 174 SPSCLEEFPRASFLCHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKII 233
DB 910 SPSCLEEFPRASFLCHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKII 969
QY 234 SRCQVCMK 241
DB 970 SRCQVCMR 977

RESULT 13
Q61436 PRELIMINARY; PRT; 253 AA.
ID Q61436
AC Q61436
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
laminae: Sequence, distribution, association with laminins, and
developmental switches."
RL J. Cell Biol. 127:879-891(1994).
DR EMBL; Z35168; CA84531.1; -.
DR F1R; I48304; I48304.
DR MGP; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match      71.1%; Score 952.5; DB 11; Length 253;
Best Local Similarity 68.7%; Pred. No. 1.5e-89;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKKGKDGSGPATWT-TRGFVTRHSQTTAIPSCPEGTVLYSGFSFLVQGNQRAHQ 59
DB 10 GPDGLQGPPEPGTTSVAHGLVTRHSQTTAIPSCPEGTVLYSGFSFLVQGNQRAHQ 69

QY 60 DLGTGSCLOQFTTNPFLFCNVNDVQNFASRNDYSYWLSTPALMPNMMAPIITGRALEPYI 119
DB 70 DLGTAGSCLRFSTMPFLFCNVNDVQNFASRNDYSYWLSTPEPMPNMMEFLKQGSIQPFI 129

QY 120 SRCVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFIFIMFTSAGSEGTGOALASPGSCL 179
DB 130 SRCVCEAPAMVIAVHSQTTQIIPCPGWSLWIGYSFVNMHTSAGAEGSQALASPGSCL 189

QY 180 EFRASPFLECHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKIIISRCQVC 239
DB 190 EFRASPFIECHGRGTCNYSNSYSFWLASINPERMPKPIPTSTVKAGELEKIIISRCQVC 249
QY 240 MKK 242
DB 250 MKR 252

RESULT 14
Q80V57 PRELIMINARY; PRT; 585 AA.
ID Q80V57

```

```

RESULT 15
Q8BNS7 PRELIMINARY; PET; 799 AA.
ID Q8BNS7;
AC Q8BNS7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
OS COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCBI_TaxID=10090;
OX (1)_TaxID=10090;
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cortex;
RC MEDLINE=22354683; PubMed=12456851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
"Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RT Nature 420:563-573(2002).
RL NMEJ; AX080682; BAC37980.1; -.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR Prodom; PD000007; Clg_helix; 1.
DR Prodom; PD003923; ProcollagnC4; 1.
DR SMART; SMO0111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 71.1%; Score 952.5; DB 11; Length 799;
Best Local Similarity 68.7%; Pred. No. 5.9e-89;
Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKRGKDGSGPATWT-TRGFVTRHSQTATPSCPEGTVPVLYSGFSFLFVQGNORAHGQ 59
DB 556 GPDGLQGPPGPGTTSVAHGFILTRHSQTTEAPQCPRGTVHIYEGFSLLYVQGNKRAHQ 615

QY 60 DLDTGLSGCLORFTYTPFLFCNVNDVCFNFRSNDYSYSLSTPALMPMNPAPITGRALSPYI 119
DB 616 DLGTAGSCLARRFTMPFMFCNNVNCVNFMRNDYSYSLSTPEPMNMNPEPKGOSIQPFI 675

QY 120 SRCVTCVSGPAIAIAVHSQTDDIPCPCHGWSLWKGPSFTIMFTSAGSEGTGQALASPQSCL 179
DB 676 SRCVCAEAPAVIAVHSQTITQIPHCPQGWDSLWIGVSFMWHTSACAEGSGQALASPQSCL 735

QY 180 EEFRASPFLECHRGTCNTYNSYSYFWLASLNPDMFRKPIPSTVKAGELSKIISRCQVC 239
DB 736 EEFRASPFIECHRGTCNTYNSYSYFWLASLNPDMFRKPIPSTVKAGELSKIISRCQVC 295

QY 240 MKK 242
DB 796 MKR 798

Search completed: April 5, 2004, 07:03:57
Job time : 149.518 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 219.777 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKKGKDGSGSPATWTRGF.....KAGELEKIISRQVCVKKRH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	244	5	ABG79219 Human Goo
2	1340	100.0	244	5	Aau75595 Human typ
3	1340	100.0	244	6	ADA20225 Human typ
4	1340	100.0	245	3	AAY67942 Human typ
5	1340	100.0	245	5	Aau75589 Human typ
6	1340	100.0	1670	7	ADD47063 Human P:O
7	1337	99.8	244	5	ABG79218 Human typ
8	1336	99.7	244	5	ABG79217 Human typ
9	1317	98.3	254	5	Aau75598 Human typ
10	1314	98.1	268	2	AAY31993 Type IV c
11	1314	98.1	268	3	AAY97555 Human alp
12	1271	94.9	232	7	ADC17697 Human typ
13	1210.5	90.3	471	2	AAY44171 Bovine ty
14	1210.5	90.3	471	3	AAY56783 Bovine al
15	1210.5	90.3	471	4	AAY56783 Bovine al
16	1199.5	89.5	471	2	AAY93483 Bovine al
17	1192	89.0	218	2	AAY44172 Partial s
18	1192	89.0	218	3	AAY44172 Human typ
19	1192	89.0	218	3	AAY56784 Human alp
20	1170	87.3	218	2	AAY93484 Human alp
21	1164	86.9	230	7	AAY93164 Partial s
22	1052	78.5	191	5	ADD47061 Rat Prote
23	1050	78.4	191	5	ADA20260 Human tum
24	960	71.6	406	3	Aau75596 Human typ
25	960	71.6	1669	4	ABG58169 Lung canc
					Aam40863 Human pol

26	960	71.6	1669	5	ABB90760 Human Tum
27	960	71.6	1669	6	ABU54467 Human tum
28	960	71.6	1672	4	AM35077 Human pol
29	951	71.0	1669	5	ABBS7334 Mouse isc
30	943.5	70.4	772	2	AAR23873 Human alp
31	943.5	70.4	772	2	AAW09643 Human typ
32	943.5	70.4	1685	4	ABG04839 Novel hum
33	943.5	70.4	1693	4	ABG15619 Novel hum
34	939	70.1	229	3	AAY67943 Human typ
35	939	70.1	229	3	AAY75587 Human typ
36	939	70.1	229	6	ADA20217 Human typ
37	939	70.1	229	7	ADC17695 Human typ
38	939	70.1	260	2	AAY31991 Type IV c
39	939	70.1	260	3	AAY97553 Human alp
40	938.5	70.0	264	2	AAY31995 Type IV c
41	938.5	70.0	264	3	AAY97557 Human alp
42	937.5	70.0	309	3	AB54044 Human pan
43	928	69.3	229	7	ADC17699 Human typ
44	923	68.9	229	1	AAP93524 Complete
45	880	65.7	211	3	AAY95918 Human Goo

## ALIGNMENTS

RESULT 1

ABG79219

ID ABG79219 standard; protein; 244 AA.

XX AC ABG79219;

XX DT 15-NOV-2002 (first entry)

XX DE Human Goodpasture disease-related protein.

XX KW Goodpasture antigen binding protein; Goodpasture syndrome;

XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;

XX KW autoimmune condition; phosphorylation; myelin basic protein; MSP;

XX KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;

XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;

XX KW pemphigoid; lichen planus.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200261430-A2.

XX PD 08-AUG-2002.

XX PF 31-JAN-2002; 2002WO-EP001010.

XX PR 31-JAN-2001; 2001US-0265249P.

XX PA (SAUS/) SAUS J.

XX PI Saus J;

XX DR WPI; 2002-619280/66.

XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.

XX PT Goodpasture syndrome or lupus, comprises identifying compounds that

XX PT reduce phosphorylation of, or formation of conformational isomers of,

XX PT target proteins.

XX PS Disclosure; Page 213-214; 217pp; English.

XX CC The invention relates to identifying candidate compounds to treat an

XX CC autoimmune condition by identifying compounds that reduce phosphorylation

XX CC of a first target protein (I) (which is selected from Goodpasture antigen

XX CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)

XX CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-

XX CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising

XX CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of



RESULT 3  
ADA20225  
ID ADA20225 standard; protein; 244 AA.  
XX AC ADA20225;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human type IV collagen alpha 3 chain partial protein sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human.  
XX OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 1..244  
FT /note= "Tumstatin"  
FT Region 1..124  
FT /note= "Tumstatin 333; pET22b-alpha 3 (IV) NCI region"  
FT Region 1..19  
FT /note= "T1 peptide"  
FT Region 28..42  
FT /note= "First Goodpasture epitope"  
FT Region 53..72  
FT /note= "T2 peptide"  
FT Region 68..88  
FT /note= "T3 peptide"  
FT Region 83..102  
FT /note= "T4 peptide"  
FT Region 98..116  
FT /note= "T5 peptide"  
FT Region 113..132  
FT /note= "T6 peptide"  
FT Region 125..244  
FT /note= "Tumstatin 334"  
FT Region 139..152  
FT /note= "Second Goodpasture epitope"  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 52; Fig 18; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
with anti-angiogenic properties. The invention also relates to the DNA  
sequences which encode the novel proteins. A wide variety of diseases are  
the result of undesirable angiogenesis. The formation of new capillaries  
from pre-existing vessels is essential for tumor growth and metastasis.  
XX CC Basement membrane organisation is dependent on the assembly of a type IV  
collagen network which may occur through the C-terminal globular non-  
collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
forms are ubiquitously exhibited in human basement membranes. In the  
present invention, cell surface receptors (in particular integrins) which  
specifically bind anti-angiogenic proteins and peptides (in particular  
the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
collagen) are disclosed. The proteins of the invention may inhibit tumour  
growth, angiogenic activity in mammalian tissue or protein synthesis in  
endothelial cells and thus may exhibit cytostatic activity. The DNA  
sequences of the invention may be useful in gene therapy. The present  
sequence is the partial amino acid sequence of the alpha 3 chain of human  
type IV collagen. The "tumstatin" protein of the invention was derived  
from this protein and comprises the full length of the present sequence.  
XX SQ Sequence 244 AA;  
Query Match 100.0%; Score 1340; DB 6; Length 244;  
Best Local Similarity 100.0%; Pred No. 2.9e-133;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
QY 61 LGTLGSCLORETTMPFLEFCNVNDYCNFASRNDYVWLSLTPALPMNMAPITGRALEPVIS 120  
DB 61 LGTLGSCLORETTMPFLEFCNVNDYCNFASRNDYVWLSLTPALPMNMAPITGRALEPVIS 120  
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALSPGSCLE 180  
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALSPGSCLE 180  
QY 181 EFRASPFLCHGRGTCNVYNSYSFSLWLASLNPFRKPISTVKAGELEKIIIRCOVCM 240  
DB 181 EFRASPFLCHGRGTCNVYNSYSFSLWLASLNPFRKPISTVKAGELEKIIIRCOVCM 240  
QY 241 KKRH 244  
DB 241 KKRH 244  
RESULT 4  
AAV67942  
ID AAV67942 standard; protein; 245 AA.  
XX AC AAV67942;  
XX DT 03-APR-2000 (first entry)  
XX DE Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.  
XX KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;  
KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;  
KW ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;  
KW myocardial angiogenesis; plaque neovascularisation; angiofibroma;  
KW atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;  
KW contraception; obesity.  
XX OS Homo sapiens.  
XX PN WO9965940-A1.  
XX PD 23-DEC-1999.  
XX PF 17-JUN-1999; 99WO-US013737.  
XX PR 17-JUN-1999; 98US-0089689P.  
XX PR 25-MAR-1999; 99US-0126175P.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2000-097708/08.  
XX DR N-PSDB; AA257158.  
XX PT Anti-angiogenic proteins comprising the NCI domain of the alpha 1, 2 or 3

PT chain of Type IV collagen used in, e.g. treatment of benign tumors and  
 XX rheumatoid arthritis.  
 PS Claim 32; Fig 16B; 117pp; English.  
 XX  
 CC The present sequence represents the human type IV collagen alpha 3 chain.  
 CC The present invention describes an isolated protein chosen from the NCI  
 CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a  
 CC fragment, analogue, derivative or mutant, which has anti-angiogenic  
 CC properties. The anti-angiogenic proteins, multimers and chimeras are  
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially  
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign  
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular  
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemophilic joints, arterio-  
 CC sclerosis, angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori  
 CC ulcers, dialysis graft vascular access stenosis, contraception and  
 CC obesity. The compositions can be used to inhibit a disease characterised  
 CC by angiogenic activity, in conjunction with radiation therapy,  
 CC chemotherapy or immunotherapy  
 XX  
 SQ Sequence 245 AA;

Query Match 100.0%; Score 1340; DB 3; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-133;  
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
 DB 2 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 61  
 QY 61 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRLPEYIS 120  
 DB 62 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRLPEYIS 121  
 QY 121 RCTVCEGPAIAVHVSQTTPDIPPCPHGWSLWKGFSPFIMFTSAGSEGTGQALSPGSCLE 180  
 DB 122 RCTVCEGPAIAVHVSQTTPDIPPCPHGWSLWKGFSPFIMFTSAGSEGTGQALSPGSCLE 181  
 QY 181 EFRASPFLCHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKAGELEKILSRQVCM 240  
 DB 182 EFRASPFLCHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKAGELEKILSRQVCM 241  
 QY 241 KKRH 244  
 DB 242 KKRH 245

RESULT 5  
 AAU75589  
 ID AAU75589 standard; protein; 245 AA.

XX AAU75589;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human type IV collagen alpha 3 chain, 'Tumstatin'.  
 XX  
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151523-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 08-JAN-2001; 2001WO-US0000565.  
 XX  
 PR 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.  
 PR 21-JUL-2000; 2000US-00625191.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Kalluri R;  
 XX  
 DR WPI; 2002-188037/24.  
 DR N-FSDB; ABK15365.  
 XX  
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.  
 XX  
 PS Claim 29; Fig 18B; 205pp; English.  
 CC  
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain  
 XX  
 SQ Sequence 245 AA;

Query Match 100.0%; Score 1340; DB 5; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-133;  
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
 DB 2 GLKRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSLFVQGNQRAHQD 61  
 QY 61 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRLPEYIS 120  
 DB 62 LGTLGSCLOQRTTTFPCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRLPEYIS 121  
 QY 121 RCTVCEGPAIAVHVSQTTPDIPPCPHGWSLWKGFSPFIMFTSAGSEGTGQALSPGSCLE 180  
 DB 122 RCTVCEGPAIAVHVSQTTPDIPPCPHGWSLWKGFSPFIMFTSAGSEGTGQALSPGSCLE 181  
 QY 181 EFRASPFLCHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKAGELEKILSRQVCM 240

Db 182 EFRASPFLECHGRGTCNYNSYSFWLASINPERMFRKPIPSTVKAGELEKIISRCQVCM 241  
Qy 241 KKRH 244  
Db 242 KKRH 245

Query Match 100.0%; Score 1340; DB 7; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 3.7e-132;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPFLYSGFSLFVQGNQRAHQD 60  
Db 1427 GLKGRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPFLYSGFSLFVQGNQRAHQD 1486

Qy 61 LGTLGSCLORFETTPFLFCNVNDVCFASRNDYSYWLSTPALPMNMWAPITGRALPEYIS 120  
Db 1487 LGTLGSCLORFETTPFLFCNVNDVCFASRNDYSYWLSTPALPMNMWAPITGRALPEYIS 1546

Qy 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKFSPFIMFTSAGSEGTGQALASPGSCLE 180  
Db 1547 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKFSPFIMFTSAGSEGTGQALASPGSCLE 1606

Qy 181 EFRASPFLECHGRGTCNYNSYSFWLASINPERMFRKPIPSTVKAGELEKIISRCQVCM 240  
Db 1607 EFRASPFLECHGRGTCNYNSYSFWLASINPERMFRKPIPSTVKAGELEKIISRCQVCM 1666

Qy 241 KKRH 244  
Db 1667 KKRH 1670

RESULT 7  
ABG79218  
ID ABG79218 standard; protein; 244 AA.  
XX  
AC ABG79218;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE Human type IV collagen NC1 domain mutant, alpha3(IV)NC1Aa9.  
XX  
KW Goodpasture antigen binding protein; Goodpasture syndrome;  
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;  
KW alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;  
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
KW pemphigoid; lichen planus; human; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FN WO200261430-A2.  
XX  
PD 08-AUG-2002.  
XX  
PF 31-JAN-2002; 2002WO-EP001010.  
XX  
PR 31-JAN-2001; 2001US-0265249P.  
XX  
PA (SAUS/) SAUS J.  
XX  
PI Saus J;  
XX  
DR WPI; 2002-619280/66.  
DR N-PSDB; ABS64503.  
XX  
PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
PT Goodpasture syndrome or lupus, comprises identifying compounds that  
PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.  
XX  
PS Claim 21; Page 212-213; 217pp; English.  
XX  
CC The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) (which is selected from Goodpasture antigen

ADD47063  
ID ADD47063 standard; protein; 1670 AA.  
XX  
AC ADD47063;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein NP\_000082, SEQ ID NO 12751.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
FN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GENO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
DR GENBANK; NP\_000082.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.



RESULT 8	
ABG79217	
ID	ABG79217 standard; protein; 244 AA.
XX	
XX	AC
XX	ABG79217;
XX	
XX	DT
XX	15-NOV-2002 (first entry)
XX	
DE	Human type IV collagen NC1 domain mutant, alpha3(IV)NCLasp9.
XX	
KW	Goodpasture antigen binding protein; Goodpasture syndrome;
KW	chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
KW	autoimmune condition; phosphorylation; myelin basic protein; MBP;
KW	alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;
KW	systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
KW	pemphigoid; lichen planus; human; mutant; mutein.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
XX	WO200261430-A2.
XX	
XX	08-AUG-2002.
XX	
PD	
XX	
PF	31-JAN-2002; 2002WO-EP001010.
XX	
PR	31-JAN-2001; 2001US-0265249P.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
 OS Homo sapiens.  
 XX  
 XX W0200151523-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 XX 08-JAN-2001; 2001WO-US000565.  
 XX  
 XX 07-JAN-2000; 2000US-00479118.  
 PR 04-APR-2000; 2000US-00543371.  
 PR 21-JUL-2000; 2000US-00625191.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 XX Kalluri R;  
 PI  
 XX  
 XX WPI; 2002-188037/24.  
 DR  
 XX  
 XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.  
 PT  
 XX  
 XX Example 36; Page; 205pp; English.  
 XX  
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-3, which consists of residues  
 CC 133-244 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75599)  
 XX  
 XX Sequence 254 AA;

Query Match

98.3%; Score 1317; DB 5; Length 254;

Best Local Similarity 96.0%; Pred. No. 8.4e-131;  
 Matches 243; Conservative 0; Mismatches 0; Indels 10; Gaps 1;  
 QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTPAIPSCPECTGPLYSGFSFLVQGNQRAHQD 60  
 DB 2 GLKGRGDSGSPATWTRGFVTRHSQTTPAIPSCPECTGPLYSGFSFLVQGNQRAHQD 61  
 QY 61 -----LGTLSGCLQRTMPFLFCNVNDVCFASNDYSYWLSTALPMNMWAPI 110  
 DB 62 LGTSGCLQRLGTLGSCLOQRTMPFLFCNVNDVCFASNDYSYWLSTALPMNMWAPI 121  
 QY 111 TGRALEPYISRCTVCEGPAIAIAVHSQTTPDIPCPHGWSLWKGFSGFIMFTSAGSEGTGQ 170  
 DB 122 TGRALEPYISRCTVCEGPAIAIAVHSQTTPDIPCPHGWSLWKGFSGFIMFTSAGSEGTGQ 181  
 QY 171 ALASFGSCLEBFRAFPFLECHGRGTCNYNSNSYFWLASLNPERMPFKPIPTSVKAGELE 230  
 DB 182 ALASFGSCLEBFRAFPFLECHGRGTCNYNSNSYFWLASLNPERMPFKPIPTSVKAGELE 241  
 QY 231 KIISRCQVCKKR 243  
 DB 242 KIISRCQVCKKR 254  
 RESULT 10  
 AAY31993  
 ID AAY31993 standard; protein; 268 AA.  
 XX AC AAY31993;  
 XX DT 05-JAN-2000 (first entry)  
 XX DE Type IV collagen NC1 domain alpha-3 monomer.  
 XX KW Type IV collagen; NC1 domain; non-collagenous domain; human;  
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
 KW rheumatoid arthritis; retinal neovascularization;  
 KW choroidal neovascularization; macular degeneration;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidermal keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; soggrens; acne rosacea; phlyctenulosis;  
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;  
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;  
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;  
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;  
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;  
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;  
 KW pemphigoid.  
 OS Homo sapiens.  
 XX Synthetic.  
 Key Location/Qualifiers  
 Peptide 1..17 /note= "EM40 signal peptide"  
 Protein 18..268 /note= "mature protein"  
 Peptide 18..25 /note= "affinity tag"  
 Protein 26..268 /note= "NC1 alpha-3 monomer"  
 WO9949885-A2.  
 PN

```

XX PD 07-OCT-1999.
XX XX
XX PF 26-MAR-1999; 99WO-US006445.
XX XX
XX PR 27-MAR-1998; 98US-0079783P.
XX PR 29-OCT-1998; 98US-0106170P.
XX XX
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX XX
XX PI Hudson BG, Sarraz MP;
XX XX
XX DR WPI: 1999-601297/51.
XX DR N-PSDB; RAZ20091.
XX XX
XX PT Inhibition of angiogenesis with non-collagenous alpha chain monomer
XX PT useful for treating e.g. tumor growth or metastasis, neovascularisation,
XX PT etc.
XX PS Disclosure; Fig 17c; 56pp; English.
XX XX
XX CC This sequence represents a recombinant type IV collagen non-collagenous
XX CC (NC1) domain alpha-3 polypeptide composed of a B40 signal sequence
XX CC (which is cleaved from the mature protein) to facilitate protein
XX CC secretion, and a mature protein comprising an affinity tag (facilitates
XX CC purification and identification of the material) and the alpha-1 chain
XX CC monomer. The invention provides methods and kits for inhibiting
XX CC angiogenesis, tumour growth and metastasis, and endothelial cell
XX CC interaction with the extracellular matrix, each method comprising
XX CC contacting the tumour or animal tissue with 1 or more isolated type IV
XX CC collagen NC1 alpha chain monomer(s) selected from the group consisting of
XX CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-
XX CC 96). The monomers can be produced via recombinant protein expression. The
XX CC polynucleotides and polypeptides are used to treat an angiogenesis-
XX CC mediated disorder or condition, especially selected from solid and blood-
XX CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
XX CC neovascularization, choroidal neovascularization, macular degeneration,
XX CC corneal neovascularization, retinopathy of prematurity, corneal graft
XX CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
XX CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
XX CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
XX CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
XX CC simplex infections, herpes zoster infections, protozoan infections,
XX CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
XX CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's
XX CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
XX CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,
XX CC vein occlusion, artery occlusion, carotid obstructive disease, chronic
XX CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Behcets
XX CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic
XX CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
XX CC complications, abnormal proliferation of fibrovascular tissue,
XX CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
XX CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
XX CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
XX XX
XX SQ Sequence 268 AA;

```

```

Query Match
Best Local Similarity 98.1%; Score 1314; DB 2; Length 268;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 5 KRSGSPATWTRGTFVTRHSQTATPSCEPGTVPLVYSGFSLFVCGNQRAGQDLGTL 64
DB 29 KRSGSPATWTRGTFVTRHSQTATPSCEPGTVPLVYSGFSLFVCGNQRAGQDLGTL 88
QY 65 GSCLRFTTMTFLFCNVNDVNCNEARNDSYVWLTSTPLMPNMWAPITGRALEPIVSCTV 124
DB 89 GSCLRFTTMTFLFCNVNDVNCNEARNDSYVWLTSTPLMPNMWAPITGRALEPIVSCTV 148
QY 125 CEGFAIAIAVHSQTDTIPPCPHGWISLWKGSFIMFTSAGSEGAGQALASPGSCLEEFRA 184

```

```

Db 149 CEGFAIAIAVHSQTDTIPPCPHGWISLWKGSFIMFTSAGSEGAGQALASPGSCLEEFRA 208
QY 185 SPFLECHGRGTCNTYNSYSFSLASLNPFRFRKPISTVTKAGSELEKIIIRCQVCMKRRH 244
DB 209 SPFLECHGRGTCNTYNSYSFSLASLNPFRFRKPISTVTKAGSELEKIIIRCQVCMKRRH 268

RESULT 11
AAY97555
ID AAY97555 standard; protein; 268 AA.
XX
XX AC AAY97555;
XX
XX DT 12-FEB-2001 (first entry)
XX
XX DE Human alpha3(IV)NC1 protein sequence.
XX
XX KW Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
XX KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
XX KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
XX KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
XX KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
XX KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
XX KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
XX KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.
XX
XX OS Homo sapiens.
XX
XX PN WO200059532-A1.
XX
XX PD 12-OCT-2000.
XX
XX PF 31-MAR-2000; 2000WO-US008678.
XX
XX PR 01-APR-1999; 99US-0127391P.
XX
XX PA (BIOS-) BIOSTRATUM INC.
XX
XX PI Brooks P, Hudson B;
XX
XX DR WPI: 2000-664962/54.
XX DR N-PSDB; AAA90993.
XX
XX PT Use of antagonists of specific integrin receptors for inhibiting
XX PT angiogenesis, tumor growth or metastases, or endothelial cell
XX PT interactions with the extracellular matrix.
XX
XX PS Disclosure; Fig 17c; 78pp; English.
XX
XX CC This sequence is a human type IV collagen alpha chain monomer, designated
XX CC alpha3(IV)NC1. The invention relates to a method for inhibiting
XX CC angiogenesis, tumor growth or metastases, or endothelial cell
XX CC interactions with the extracellular matrix, comprising contacting the
XX CC cells or tissue with a polypeptide composition containing antagonists of
XX CC specific integrin receptors. The methods and the antagonists are useful
XX CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
XX CC cell interaction with the extracellular matrix. The antagonists are also
XX CC useful for treating diseases and conditions with accompanying undesired
XX CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
XX CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
XX CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
XX CC treating non-tumorigenic diseases and conditions with accompanying
XX CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
XX CC retinal neovascularisation, choroidal neovascularisation, muscular
XX CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
XX CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
XX CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
XX CC complications, chronic inflammation or psoriasis
XX
XX SQ Sequence 268 AA;

```

```

Query Match
Best Local Similarity 98.1%; Score 1314; DB 3; Length 268;
99.6%; Pred. No. 1.9e-130;

```

Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGSQSPATWTRGVFTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTL 64  
 DB 29 KRGSQSPATWTRGVFTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTL 88  
 QY 65 GSCILQRTTTPFFLCNVNDVNCNPNASNDYSYWLSTPALMPMNPAPITGRALBPYISRCTV 124  
 DB 89 GSCILQRTTTPFFLCNVNDVNCNPNASNDYSYWLSTPALMPMNPAPITGRALBPYISRCTV 148  
 QY 125 CEGPAIAIVHSQTTPDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEBPRA 184  
 DB 149 CEGPAIAIVHSQTTPDIPPCPHGWISLWKGFSTIMFTSAGSEGTGQALASPGSCLEBPRA 208  
 QY 185 SPFLCHGRGTCNYSNSYSFNLASLNPFRMFKPIPTVKAGELEKIISRCQVCMKKRH 244  
 DB 209 SPFLCHGRGTCNYSNSYSFNLASLNPFRMFKPIPTVKAGELEKIISRCQVCMKKRH 268

RESULT 12  
 ADC17697  
 ID ADC17697 standard; protein; 232 AA.  
 AC ADC17697;  
 DT 18-DEC-2003 (first entry)  
 DE Human type IV collagen alpha 3 chain protein SEQ ID NO:304.  
 KW crystallised NCI domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 OS Homo sapiens.  
 XX WO2003012122-A2.  
 PN 13-FEB-2003.  
 XX 26-JUL-2002; 2002WO-US023763.  
 XX 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX Sundaramoorthy M, Hudson B;  
 PI WPI; 2003-332730/31.  
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX Disclosure; SEQ ID NO 304; 169pp; English.  
 PS The present invention describes a crystallised NCI domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the

CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NCI  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents an amino acid sequence which is used in the exemplification of  
 CC the present invention.

XX Sequence 232 AA;

Query Match 94.9%; Score 1271; DB 7; Length 232;  
 Best Local Similarity 99.6%; Pred. No. 5.6e-126;  
 Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 ATWTRGFVTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTLGSLQRT 72  
 DB 1 ATWTRGFVTRHSQTTPSCPEGTPLVYSGFSLFVQGNQRAHQDGLGTLGSLQRT 60  
 QY 73 TNPFLFCNVNDVNCNPNASNDYSYWLSTPALMPMNPAPITGRALBPYISRCTVCGPAIAI 132  
 DB 61 TNPFLFCNVNDVNCNPNASNDYSYWLSTPALMPMNPAPITGRALBPYISRCTVCGPAIAI 120  
 QY 133 AVHSQTTDIPCPHGWISLWKGFSTIMFTSAGSGTGOALASPGSCLEBPRAFPFLSCHG 192  
 DB 121 AVHSQTTDIPCPHGWISLWKGFSTIMFTSAGSGTGOALASPGSCLEBPRAFPFLSCHG 180  
 QY 193 RGTGNYNSYSFNLASLNPFRMFKPIPTVKAGELEKIISRCQVCMKKRH 244  
 DB 181 RGTGNYNSYSFNLASLNPFRMFKPIPTVKAGELEKIISRCQVCMKKRH 232

RESULT 13  
 AAY44171  
 ID AAY44171 standard; protein; 471 AA.  
 XX AAY44171;  
 AC AAY44171;  
 DT 01-FEB-2000 (first entry)  
 XX Bovine type IV collagen alpha3 chain protein.  
 DE Recombinant; bovine; alpha3 chain; type IV collagen; detection;  
 KW Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.  
 XX Bos taurus.  
 OS US5973120-A.  
 PN 26-OCT-1999.  
 PD 07-MAR-1995; 95US-00399889.  
 XX 30-NOV-1990; 90US-00621091.  
 PR (UYVA ) UNIV YALE.  
 PA (UNIV ) UNIV KANSAS MEDICAL CENT.  
 XX Hudson BG, Reiders ST, Morrison KE;  
 PI WPI; 1999-610317/52.  
 DR N-PSDB; AA228774.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for  
 PT diagnosis and treatment of Goodpasture syndrome.  
 XX  
 XX Claim 1; Col 31-34; 27pp; English.  
 XX  
 CC This sequence represents a recombinant bovine alpha3 chain of type IV  
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids of  
 CC the C-terminal end of the triple helical domain and all 233 amino acids  
 CC of the C-terminal non-collagenous domain. Alpha3 chain collagen  
 CC polypeptides are useful for detecting Goodpasture antibodies in blood or  
 CC tissue from a human patient and for treating Goodpasture syndrome,  
 CC especially by neutralising the antibodies in the blood. The polypeptides  
 CC also have a nephrotrophic activity  
 XX  
 SQ Sequence 471 AA;

Query Match 90.3%; Score 1210.5; DB 2; Length 471;  
 Best Local Similarity 90.6%; Pred. No. 3.6e-119;  
 Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;  
 QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPAIPSCPGTVPVLYSGFSFLVQGNQAHGQ 59  
 DB 227 GLKGRPGDTGPPAAGAVNRGFFVTRHSQTTPAIPSCPGTVPVLYSGFSFLVQGNQAHGQ 286  
 QY 60 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFRNDYSYWLSTPALMPMNPAPITGRALEPYI 119  
 DB 287 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFRNDYSYWLSTPALMPMNPAPITGRALEPYI 346  
 QY 120 SRTVCCEGPAIAIVHSQTTDIPCPHGWISLWKGFIFMFTSAGSEGTQALASPGSCL 179  
 DB 347 SRTVCCEGPAIAIVHSQTTDIPCPHGWISLWKGFIFMFTSAGSEGTQALASPGSCL 406  
 QY 180 EEFRAFPFIECHGRGTCNYNSYSFWLASLNPERMPKPIPTSVKAGELEKIISRCQVC 239  
 DB 407 EEFRAFPFIECHGRGTCNYNSYSFWLASLNPERMPKPIPTSVKAGELEKIISRCQVC 466  
 QY 240 MKKR 243  
 DB 467 MKMR 470

RESULT 14  
 AAY56783  
 ID AAY56783 standard; protein; 471 AA.  
 AC AAY56783;

XX 27-MAR-2000 (first entry)  
 XX Bovine alpha3 type IV collagen.  
 XX Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.  
 XX Bos sp.

XX US6007980-A.  
 XX 28-DEC-1999.  
 XX 07-OCT-1998; 98US-00167364.  
 XX 30-NOV-1990; 90US-00621091.  
 XX 07-MAR-1995; 95US-00399889.  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 XX (UYVA ) UNIV YALE.

XX Hudson BG, Reenders ST, Morrison KE;  
 XX WPI; 2000-096371/08.  
 XX N-PSDB; AAZ46728.

PT Diagnosing and treating Goodpasture syndrome using a peptide derived from  
 PT type IV collagen.  
 XX  
 XX Disclosure; Col 19-24; 26pp; English.  
 XX  
 CC The invention provides a method of detecting Goodpasture antibodies in  
 CC the fluid of a patient by contacting it with a peptide comprising at most  
 CC 218 amino acids of the human alpha3 chain type IV collagen that contains  
 CC the fragment shown in AAY56783. The methods are useful for the diagnosis  
 CC and treatment of Goodpasture syndrome. The present sequence represents  
 CC the bovine alpha3 chain of type IV collagen  
 XX  
 SQ Sequence 471 AA;

Query Match 90.3%; Score 1210.5; DB 3; Length 471;  
 Best Local Similarity 90.6%; Pred. No. 3.6e-119;  
 Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;  
 QY 1 GLKGRGDSGSPATWTT-RGFVTRHSQTTPAIPSCPGTVPVLYSGFSFLVQGNQAHGQ 59  
 DB 227 GLKGRPGDTGPPAAGAVNRGFFVTRHSQTTPAIPSCPGTVPVLYSGFSFLVQGNQAHGQ 286  
 QY 60 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFRNDYSYWLSTPALMPMNPAPITGRALEPYI 119  
 DB 287 DLGTLGSCLOQRTTTPFLFCNVNDVCFNFRNDYSYWLSTPALMPMNPAPITGRALEPYI 346  
 QY 120 SRTVCCEGPAIAIVHSQTTDIPCPHGWISLWKGFIFMFTSAGSEGTQALASPGSCL 179  
 DB 347 SRTVCCEGPAIAIVHSQTTDIPCPHGWISLWKGFIFMFTSAGSEGTQALASPGSCL 406  
 QY 180 EEFRAFPFIECHGRGTCNYNSYSFWLASLNPERMPKPIPTSVKAGELEKIISRCQVC 239  
 DB 407 EEFRAFPFIECHGRGTCNYNSYSFWLASLNPERMPKPIPTSVKAGELEKIISRCQVC 466  
 QY 240 MKKR 243  
 DB 467 MKMR 470

RESULT 15  
 AAE09483  
 ID AAE09483 standard; protein; 471 AA.  
 AC AAE09483;

XX 19-NOV-2001 (first entry)  
 XX Bovine alpha-3 chain of type IV collagen protein.  
 XX Bovine; alpha-3 chain; type IV collagen; immunosuppressive; therapy;  
 XX Goodpasture syndrome.

XX Bos taurus.  
 XX US6277558-B1.  
 XX 21-AUG-2001.  
 XX 12-NOV-1999; 99US-00439897.  
 XX 30-NOV-1990; 90US-00621091.  
 XX 07-MAR-1995; 95US-00399889.  
 XX 07-OCT-1998; 98US-00167364.  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG;  
 XX WPI; 2001-540401/50.  
 XX N-PSDB; AAD16399.

XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting  
 PT Goodpasture antibodies from bodily fluid/tissue from patient or for



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 153.017 Seconds

(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 1340

Sequence: 1 GLKGRGDSGPATWTRGF.....KAGELEKIISRCQVCVKKEH 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	244	14	US-10-032-221B-10
2	1271	94.9	232	14	US-10-032-221B-10
3	1052	78.5	191	14	US-10-032-221B-22
4	960	71.6	406	9	US-09-925-302-507
5	960	71.6	1669	15	US-10-372-683-8
6	939	70.1	229	14	US-10-032-221B-2
7	939	70.1	229	14	US-10-032-221B-2
8	937.5	70.0	309	9	US-09-925-297-496
9	928	69.3	229	14	US-10-032-221B-306
10	880	65.7	211	14	US-10-270-877-46
11	880	65.7	211	14	US-10-270-877-46
12	852.5	63.6	1744	15	US-10-369-493-5832
13	783.5	58.5	1759	15	US-10-369-493-7032
14	755.5	56.4	430	9	US-09-925-302-518
15	755.5	56.4	459	15	US-10-331-496A-27

16	755.5	56.4	459	15	US-10-372-683-30	Sequence 30, Appl
17	755.5	56.4	1712	10	US-09-861-403-9	Sequence 9, Appl
18	741.5	55.3	227	14	US-10-032-221B-6	Sequence 303, Appl
19	741.5	55.3	227	14	US-10-032-221B-6	Sequence 6, Appl
20	727	54.3	228	14	US-10-032-221B-307	Sequence 307, Appl
21	721	53.8	132	14	US-10-032-221B-23	Sequence 23, Appl
22	690.5	51.5	231	14	US-10-032-221B-305	Sequence 305, Appl
23	678	50.6	124	14	US-10-032-221B-20	Sequence 20, Appl
24	662	49.4	120	14	US-10-032-221B-21	Sequence 21, Appl
25	619	46.2	112	14	US-10-032-221B-33	Sequence 33, Appl
26	480	35.8	88	14	US-10-032-221B-33	Sequence 34, Appl
27	472.5	35.3	143	14	US-10-032-221B-34	Sequence 34, Appl
28	471	35.1	88	14	US-10-032-221B-25	Sequence 25, Appl
29	433	32.3	79	14	US-10-032-221B-26	Sequence 26, Appl
30	353	26.3	64	14	US-10-032-221B-25	Sequence 25, Appl
31	337	25.1	142	9	US-09-864-761-38021	Sequence 38021, A
32	334	24.9	68	14	US-10-270-877-50	Sequence 50, Appl
33	334	24.9	68	14	US-10-270-877-50	Sequence 50, Appl
34	334	24.9	72	14	US-10-270-877-48	Sequence 48, Appl
35	334	24.9	72	14	US-10-270-877-52	Sequence 52, Appl
36	334	24.9	72	14	US-10-270-877-52	Sequence 48, Appl
37	334	24.9	72	14	US-10-270-877-52	Sequence 52, Appl
38	328	24.5	72	14	US-10-270-877-61	Sequence 61, Appl
39	328	24.5	72	14	US-10-270-877-61	Sequence 61, Appl
40	260	19.4	70	11	US-09-864-408A-1258	Sequence 1258, A
41	201	15.0	46	9	US-09-864-761-48095	Sequence 48095, A
42	189.5	14.1	70	9	US-09-864-761-37448	Sequence 37448, A
43	189.5	14.1	70	9	US-09-864-761-47938	Sequence 47938, A
44	147	11.0	25	14	US-10-032-221B-37	Sequence 37, Appl
45	146	10.9	27	14	US-10-032-221B-39	Sequence 39, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-032-221B-10  
; Sequence 10, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-10

Query Match 100.0%; Score 1340; DB 14; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.1e-132; Indels 0; Gaps 0;  
Matches 244; Conservative 0; Mismatches 0;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTATPSCSEGVPLYSFPLVQGNRAHQD 60



Db 1 GLKRGDSGSDATWTTTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQD 60  
QY 61 LGTLGSLQRFRTMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGRALPEYIS 120  
Db 61 LGTLGSLQRFRTMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGRALPEYIS 120  
QY 121 RCTVCEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180  
Db 121 RCTVCEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLE 180  
QY 181 EFRASPFLCHGRGTCNVYNSYSFNLASLNPERMFKPIPTSVKAGELEKIIISRCQVCM 240  
Db 181 EFRASPFLCHGRGTCNVYNSYSFNLASLNPERMFKPIPTSVKAGELEKIIISRCQVCM 240  
QY 241 KKRH 244  
Db 241 KKRH 244

## RESULT 2

US-10-206-699-304  
; Sequence 304, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MEHB 01-1017  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US/10/206,699  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 304  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 3 chain  
US-10-206-699-304

Query Match 94.9%; Score 1271; DB 14; Length 232;  
Best Local Similarity 99.6%; Pred. No. 2e-125;  
Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 13 ATWTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLSCLQRF 72  
Db 1 ATWTRGVFTRHSQTTAIPSCPEGTVPVLYSGFSLFVQGNQRAHGQDGLTGLSCLQRF 60  
QY 73 TMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGRALPEYISRCTVCEGPAIAI 132  
Db 61 TMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGRALPEYISRCTVCEGPAIAI 120  
QY 133 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLEEFPRASPFLECHG 192  
Db 121 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALASPGSCLEEFPRASPFLECHG 180  
QY 193 RGTCTNYNSYSFNLASLNPERMFKPIPTSVKAGELEKIIISRCQVCMKKRH 244  
Db 181 RGTCTNYNSYSFNLASLNPERMFKPIPTSVKAGELEKIIISRCQVCMKKRH 232

## RESULT 3

US-10-032-221B-22  
; Sequence 22, Application US/10032221B

Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)  
US-10-032-221B-22

Query Match 78.5%; Score 1052; DB 14; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.8e-102;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 54 QRAHGQDGLTGLSCLQRFRTMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGR 113  
Db 1 QRAHGQDGLTGLSCLQRFRTMPFLFCNVNDVCFNPNASRNDYSYWLSTPALMPMNNAPITGR 60  
QY 114 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALA 173  
Db 61 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSGTGQALA 120  
QY 174 SPGSCLEEFPRASPFLECHGRGTCNVYNSYSFNLASLNPERMFKPIPTSVKAGELEKII 233  
Db 121 SPGSCLEEFPRASPFLECHGRGTCNVYNSYSFNLASLNPERMFKPIPTSVKAGELEKII 180  
QY 234 SRCQVCMKKRH 244  
Db 181 SRCQVCMKKRH 191

## RESULT 4

US-09-925-302-507  
; Sequence 507, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 507  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

NAME/KEY: SITE  
LOCATION: (71)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-507

Query Match 71.6%; Score 960; DB 9; Length 406;  
Best Local Similarity 69.0%; Pred. No. 2.4e-92;  
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGVFTRHSQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQD 60  
DB 166 GLPGSMGPPGTPS--VDHGLFVTRHSQTIDDPQCPGSGTKILYHGYSLVYQGNRAHQD 223  
QY 61 LGTLGSLCLOFTTTPFLFCNVNDVCFNNDYSLWSTPALMPNMNAPITGRALPEYIS 120  
DB 224 LGTAGSLCLRFSTMPFLFCNVNVCNFCNNDYSLWSTPEPMPNMAPITGENTIRPFIS 283  
QY 121 RCTVCEGPAIAIAVHSQTIDPPCPHGWISLWKGFSFIMFTSAGSEGTQOALASPGSCLE 180  
DB 284 RCACEAPAMVMAVHSQTIDPPCPGWSLWIGYFVHVSAGAGSGOALASPGSCLE 343  
QY 181 EFRASPLECHGRGTCNYNSYNSYFSLASLNPERMFKPIPTSVKAGELEKIISRCQVCM 240  
DB 344 EFRASPLECHGRGTCNYANYSFWLATIERSEMFKKPTPTLTKAGELRTHVSRQVCM 403  
QY 241 KK 242  
DB 404 RR 405

RESULT 5  
US-10-372-683-8  
Sequence 8, Application US/10372683  
Publication No. US20040009171A1  
GENERAL INFORMATION:  
APPLICANT: GERRITSEN, MARY E.  
APPLICANT: PEARLE JR., FRANKLIN V.  
TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA  
FILE REFERENCE: P1928R1P1  
CURRENT APPLICATION NUMBER: US/10/372,683  
CURRENT FILING DATE: 2003-02-21  
PRIOR APPLICATION NUMBER: US 10/271,690  
PRIOR FILING DATE: 2002-10-16  
PRIOR APPLICATION NUMBER: US 60/344,534  
PRIOR FILING DATE: 2001-10-18  
NUMBER OF SEQ ID NOS: 49  
SEQ ID NO 8  
LENGTH: 1669  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-372-683-8

Query Match 71.6%; Score 960; DB 15; Length 1669;  
Best Local Similarity 69.0%; Pred. No. 1.5e-91;  
Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

QY 1 GLKGRGDSGPATWTRGVFTRHSQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQD 60  
DB 1429 GLPGSMGPPGTPS--VDHGLFVTRHSQTIDDPQCPGSGTKILYHGYSLVYQGNRAHQD 1486  
QY 61 LGTLGSLCLOFTTTPFLFCNVNDVCFNNDYSLWSTPALMPNMNAPITGRALPEYIS 120  
DB 1487 LGTAGSLCLRFSTMPFLFCNVNVCNFCNNDYSLWSTPEPMPNMAPITGENTIRPFIS 1546  
QY 121 RCTVCEGPAIAIAVHSQTIDPPCPHGWISLWKGFSFIMFTSAGSEGTQOALASPGSCLE 180  
DB 1547 RCACEAPAMVMAVHSQTIDPPCPGWSLWIGYFVHVSAGAGSGOALASPGSCLE 1606  
QY 181 EFRASPLECHGRGTCNYNSYNSYFSLASLNPERMFKPIPTSVKAGELEKIISRCQVCM 240  
DB 1607 EFRASPLECHGRGTCNYANYSFWLATIERSEMFKKPTPTLTKAGELRTHVSRQVCM 1666

QY 241 KK 242  
DB 1667 RR 1668

RESULT 6  
US-10-206-699-302  
Sequence 302, Application US/10206699  
Publication No. US20030100510A1  
GENERAL INFORMATION:  
APPLICANT: Sundaramoorthy, M.  
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
FILE REFERENCE: MBHB 01-1017  
CURRENT APPLICATION NUMBER: US/10/206,699  
CURRENT FILING DATE: 2002-07-26  
PRIOR APPLICATION NUMBER: US 60/308,523  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: US 60/351,289  
PRIOR FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: US 60/366,854  
PRIOR FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: US 60/385,362  
PRIOR FILING DATE: 2002-06-03  
NUMBER OF SEQ ID NOS: 307  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 302  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
OTHER INFORMATION: alpha 1 chain  
US-10-206-699-302

Query Match 70.1%; Score 939; DB 14; Length 229;  
Best Local Similarity 71.9%; Pred. No. 1.9e-90;  
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTPAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLQRFTTTPPLF 78  
DB 5 GFLVTRHSQTIDDPQCPGSGTKILYHGYSLVYQGNRAHQDGLTAGSCLRKFTSTPPLF 64  
QY 79 CNVNDVCFNNDYSLWSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQT 138  
DB 65 CNVNDVCFNNDYSLWSTPEPMPNMAPITGENTIRPFISRCVACEAPAMVMAVHSQT 124  
QY 139 TDIPCPHGWISLWKGFSFIMFTSAGSEGTQOALASPGSCLEFRASPLECHGRGTCNY 198  
DB 125 IQIPCPGWSLWIGYFVHVSAGAGSGOALASPGSCLEFRASPLECHGRGTCNY 184  
QY 199 YSNSYFSLASLNPERMFKPIPTSVKAGELEKIISRCQVCMKK 242  
DB 185 YANYSFWLATIERSEMFKKPTPTLTKAGELRTHVSRQVCMRR 228

RESULT 7  
US-10-032-221B-2  
Sequence 2, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118

;; PRIOR FILING DATE: 2000-01-07  
;; PRIOR APPLICATION NUMBER: US 09/335,224  
;; PRIOR FILING DATE: 1999-06-17  
;; PRIOR APPLICATION NUMBER: US 60/126,175  
;; PRIOR FILING DATE: 1999-03-25  
;; PRIOR APPLICATION NUMBER: US 60/089,689  
;; PRIOR FILING DATE: 1998-06-17  
;; NUMBER OF SEQ ID NOS: 58  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 2  
;; LENGTH: 229  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-032-221B-2

Query Match 70.1%; Score 939; DB 14; Length 229;  
Best Local Similarity 71.9%; Pred. No. 1.9e-90;  
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;  
QY 19 GFVETRHSQTTPSCPEGTVPLYSGLFVQGNORAHGQDLGTGLGSLQRFTHMPFLF 78  
DB 5 GFLVTRHSQTIDDPQSPGSKILYHGYSLLYVQGNERAGQDLGTAGSLRKFSTWPLF 64  
QY 79 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQT 138  
DB 65 CNINNVCFASRNDYSYWLSTPEPMPMSAPITGENIRPFISRCVCEAPAMVMAVHSQT 124  
QY 139 TDTPCPHGWLISLWKGSFTMTSAGSEGTGOALASPGSLCEFRASPELECHGRGTCNY 198  
DB 125 IQTPPCFSGWSSLIWIGYFVWHTSAGAGSGQALASPGSLCEFRASPELECHGRGTCNY 184  
QY 199 YNSYSFSLASLNPFRMRKPIPTSTVKAGELEKIIISRCQVCMKK 242  
DB 185 YANYSFWLATIERSEWFKKPTSTLKAGELRTHVSRQVCMKR 228

RESULT 8  
US-09-925-297-496  
;; Sequence 496, Application US/09925297  
;; Patent No. US20020081659A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA105  
;; CURRENT APPLICATION NUMBER: US/09/925,297  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05989  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 928  
;; SOFTWARE: Patent in Ver. 2.0  
;; SEQ ID NO 496  
;; LENGTH: 309  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (247)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-496

Query Match 70.0%; Score 937.5; DB 9; Length 309;  
Best Local Similarity 67.5%; Pred. No. 3.9e-90;  
Matches 164; Conservative 35; Mismatches 43; Indels 1; Gaps 1;  
QY 1 GLKGRDGSQSPATWT-TRGFVTRHSQTTPSCPEGTVPLYSGLFVQGNORAHQ 59  
DB 66 GPDGLQFPGPPTSSVAHGFLLTRHSQTDAQCQGTQVTEGSLLYVQGNKRAHQ 125  
QY 60 DLGTGLGSLQRFTHMPFLCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYI 119  
DB 126 DLGTAGSLRKFSTMPFMCNINNVCFASRNDYSYWLSTPEPMPMSQPLKQSQTPPI 185

QY 120 SRTVCEGPAIAIAVHSQTIDTPCPHGWLISLWKGSFTMTSAGSEGTGOALASPGSL 179  
DB 186 SRCVCEAPAMVMAVHSQTIDTPCPHGQWDSLIWIGYFVWHTSAGAGSGQALASPGSL 245  
QY 180 EEFRAFPFLECHGRGTCNYNSYSFSLASLNPFRMRKPIPTSTVKAGELEKIIISRCQVC 239  
DB 246 EEFRAFPFLECHGRGTCNYNSYSFSLATVDVSDMFSKQSETLKAGDLRTRISRCQVC 305  
QY 240 MKK 242  
DB 306 MKR 308

RESULT 9  
US-10-206-699-306  
;; Sequence 306, Application US/10206699  
;; Publication No. US20030100510A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sundaramoorthy, M.  
;; APPLICANT: Hudson, B.  
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
;; FILE REFERENCE: MHB 01-1017  
;; CURRENT APPLICATION NUMBER: US/10/206,699  
;; CURRENT FILING DATE: 2002-07-26  
;; PRIOR APPLICATION NUMBER: US 60/308,523  
;; PRIOR FILING DATE: 2001-07-27  
;; PRIOR APPLICATION NUMBER: US 60/351,289  
;; PRIOR FILING DATE: 2001-10-29  
;; PRIOR APPLICATION NUMBER: US 60/366,854  
;; PRIOR FILING DATE: 2002-03-22  
;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 306  
;; LENGTH: 229  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; OTHER INFORMATION: alpha 5 chain  
US-10-206-699-306

Query Match 69.3%; Score 928; DB 14; Length 229;  
Best Local Similarity 71.0%; Pred. No. 2.7e-89;  
Matches 159; Conservative 33; Mismatches 32; Indels 0; Gaps 0;  
QY 19 GFVETRHSQTTPSCPEGTVPLYSGLFVQGNORAHGQDLGTGLGSLQRFTHMPFLF 78  
DB 5 GFLVTRHSQTIDDPQSPGSKILYHGYSLLYVQGNERAGQDLGTAGSLRKFSTWPLF 64  
QY 79 CNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYISRCTVCEGPAIAIAVHSQT 138  
DB 65 CNINNVCFASRNDYSYWLSTPEPMPMSQPLKQSQTPPIISRCVCEAPAMVMAVHSQT 124  
QY 139 TDTPCPHGWLISLWKGSFTMTSAGSEGTGOALASPGSLCEFRASPELECHGRGTCNY 198  
DB 125 IQTPPCFSGWSSLIWIGYFVWHTSAGAGSGQALASPGSLCEFRASPELECHGRGTCNY 184  
QY 199 YNSYSFSLASLNPFRMRKPIPTSTVKAGELEKIIISRCQVCMKK 242  
DB 185 YANYSFWLATVDVSDMFSKQSETLKAGDLRTRISRCQVCMKR 228

RESULT 10  
US-10-270-877-46  
;; Sequence 46, Application US/10270877  
;; Publication No. US20030049791A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Saus, Juan  
;; TITLE OF INVENTION: Goodpasture Binding Protein  
;; FILE REFERENCE: 98-723-AD1

CURRENT APPLICATION NUMBER: US/10/270,877  
CURRENT FILING DATE: 2002-10-11  
PRIOR APPLICATION NUMBER: 09/512,563  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/121,483  
PRIOR FILING DATE: 1999-02-24  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 46  
LENGTH: 211  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-877-46

Query Match 65.7%; Score 880; DB 14; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.7e-84;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHQD 60  
DB 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHQD 60  
QY 61 LGTLGSCLOFRTTTPFPCNVNDVCFNFSRNDYSYWLSTPALMPMNAIPITGRALEPYIS 120  
DB 61 LGTLGSCLOFRTTTPFPCNVNDVCFNFSRNDYSYWLSTPALMPMNAIPITGRALEPYIS 120  
QY 121 RCTVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIM 159  
DB 121 RCTVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIM 159

RESULT 11  
US-10-270-837-46  
Sequence 46, Application US/10270837  
Publication No. US20030054488A1  
GENERAL INFORMATION:  
APPLICANT: Saus, Juan  
TITLE OF INVENTION: Goodpasture Binding Protein  
FILE REFERENCE: 98-723-AD2  
CURRENT APPLICATION NUMBER: US/10/270,837  
CURRENT FILING DATE: 2002-10-11  
PRIOR APPLICATION NUMBER: 09/512,563  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/121,483  
PRIOR FILING DATE: 1999-02-24  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 46  
LENGTH: 211  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-837-46

Query Match 65.7%; Score 880; DB 14; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.7e-84;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHQD 60  
DB 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHQD 60  
QY 61 LGTLGSCLOFRTTTPFPCNVNDVCFNFSRNDYSYWLSTPALMPMNAIPITGRALEPYIS 120  
DB 61 LGTLGSCLOFRTTTPFPCNVNDVCFNFSRNDYSYWLSTPALMPMNAIPITGRALEPYIS 120  
QY 121 RCTVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIM 159  
DB 121 RCTVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIM 159

RESULT 12  
US-10-369-493-5832  
Sequence 5832, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 5832  
LENGTH: 1744  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5832

Query Match 63.6%; Score 852.5; DB 15; Length 1744;  
Best Local Similarity 60.7%; Pred. No. 3.4e-80;  
Matches 148; Conservative 35; Mismatches 60; Indels 1; Gaps 1;  
QY 1 GLKGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHQ 59  
DB 1501 GLPGTGYGSPGAWAPSRGFTTAKHSQTTAVPQCPGASQLWEGYSLLYVQNGRASGQ 1560  
QY 60 DLGLGSCLOFRTTTPFPCNVNDVCFNFSRNDYSYWLSTPALMPMNAIPITGRALEPYI 119  
DB 1561 DLGQPGSCLSKFNTMPFPCNVNSVCHVSRNDYSYWLSTDEPTPMKRPVTTAIPYI 1620  
QY 120 SRCVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIMFTSAGSEGTQALASPGSCL 179  
DB 1621 SRCVCEGPAIAVAHSQTTDIPPCPHGWISLWKGFSPFIMFTSAGSEGTQALASPGSCL 1680  
QY 180 EEFRAFPFIECHGRGTCNYNSYSYFWLASLNPERMFKPIPTVKAGELEKILSRQVC 239  
DB 1681 EEFRAFPFIECHGRGTCNYNSYSYFWLASLNPERMFKPIPTVKAGELEKILSRQVC 239  
QY 240 MKXR 243  
DB 1741 LKXR 1744

RESULT 13  
US-10-369-493-7032  
Sequence 7032, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 7032  
LENGTH: 1759  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-7032

Query Match 58.5%; Score 783.5; DB 15; Length 1759;  
Best Local Similarity 55.3%; Pred. No. 6.4e-71;  
Matches 142; Conservative 41; Mismatches 59; Indels 15; Gaps 4;  
  
QY 1 GLKGRGDSGSP-----ATWTRGVFVTRHSQTTAIPSCPEGTVPLVYSGFSLFVQ 51  
DB 1505 GLDQGGGAGLPGAPCAAGAPARDGVLVKHSQTTVEPRCPGQTKLWDGYSLLVIE 1564  
  
QY 52 GQRAHQDLGLTGLSCLOQFTTMPELFCNNVNDVCFNARNDSYVWLSLTPALMPMNAIPIT 111  
DB 1565 GNEKSHNQLGHAGSCLOQFTTMPELFCDFNNVNCYASRNDKSYWLSLTPALMPMNAIPIT 1622  
  
QY 112 GRALEYISRCTVCEGPAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQA 171  
DB 1623 EREIEYISRCVCAVECAPANTIAVHSQTTIQPCPAGWSSLWIGYFAMHTGAGAGGQGS 1682  
  
QY 172 LASPGSCLEEFRAFPLECHG-RGTCNNYSNSYFWLASLNPERMFRKPIPTVKGABLE 230  
DB 1683 LSSPGSCLEDFRATPIECNGARGSGCHYFANKFSFWLTTIDNDSFKVPESQTLKSGNLR 1742  
  
QY 231 KIISRCQVCMK---RH 244  
DB 1743 TRVSRQVCMKSTDRH 1759

RESULT 14  
US-09-925-302-518  
; Sequence 518, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: ROSEN ET AL.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 518  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (11)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-518

Query Match 56.4%; Score 755.5; DB 9; Length 430;  
Best Local Similarity 58.0%; Pred. No. 9.1e-71;  
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;  
  
QY 1 GLKGRGDSGSPATWTRGVFVTRHSQTTAIPSCPEGTVPLVYSGFSLFVQNCRAHQD 60  
DB 189 GRPGSGLPNGPGRSVSIGVLLVKHSQTDQEPXCPVGMNKLWSGYSLLYFEGOEKAHQD 248  
  
QY 61 LGTGLSCLOQFTTMPELFCNNVNDVCFNARNDSYVWLSLTPALMPMNAIPITGRALEPYIS 120  
DB 249 LGLAGSCLARFTTMPELFCNNVNDVCFNARNDSYVWLSLTPA--PLPMPVAEDEIKPYIS 306  
  
QY 121 RCTVCEGPAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQAALASPGSCLE 180  
DB 307 RCSVCEAPAIATAVHSQDVSIHPCPAGWSSLWIGYFAMHTGAGAGGQGSLSVPSGSCLE 366  
  
QY 181 EFRASPFLECH-RGTCNNYSNSYFWLASLNPERMFR-KPIPTVKGABLEKIIISRCOV 238  
DB 367 DFRATPIECNGRGRTCHYANKYSFWLTTI-PEQSQGSPSADTLKAGLIRTHISRCOV 425  
  
QY 239 CMK 241

DB 426 CMK 428  
  
RESULT 15  
US-10-331-496A-27  
; Sequence 27, Application US/10331496A  
; Publication No. US20030228305A1  
; GENERAL INFORMATION:  
; APPLICANT: FRANTZ, GRETCHEN  
; APPLICANT: HILLAN, KENNETH J.  
; APPLICANT: PHILLIPS, HEIDI S.  
; APPLICANT: POLAKIS, PAUL  
; APPLICANT: SMITH, VICTORIA  
; APPLICANT: SPENCER, SUSAN D.  
; APPLICANT: WILLIAMS, P. MICKEY  
; APPLICANT: WU, THOMAS D.  
; APPLICANT: ZHANG, ZEMIN  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF TUMOR  
; FILE REFERENCE: P5014R1-PCT  
; CURRENT APPLICATION NUMBER: US/10/331,496A  
; CURRENT FILING DATE: 2002-12-30  
; PRIOR APPLICATION NUMBER: US 60/345,444  
; PRIOR FILING DATE: 2002-01-02  
; PRIOR APPLICATION NUMBER: US 60/351,885  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: US 60/360,066  
; PRIOR FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: US 60/362,004  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/366,869  
; PRIOR FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: US 60/366,284  
; PRIOR FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 60/368,679  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 60/404,809  
; PRIOR FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: US 60/405,645  
; PRIOR FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 95  
; SEQ ID NO 27  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-331-496A-27

Query Match 56.4%; Score 755.5; DB 15; Length 459;  
Best Local Similarity 58.0%; Pred. No. 1e-70;  
Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;  
  
QY 1 GLKGRGDSGSPATWTRGVFVTRHSQTTAIPSCPEGTVPLVYSGFSLFVQNCRAHQD 60  
DB 218 GRPGSGLPNGPGRSVSIGVLLVKHSQTDQEPXCPVGMNKLWSGYSLLYFEGOEKAHQD 277  
  
QY 61 LGTGLSCLOQFTTMPELFCNNVNDVCFNARNDSYVWLSLTPALMPMNAIPITGRALEPYIS 120  
DB 278 LGLAGSCLARFTTMPELFCNNVNDVCFNARNDSYVWLSLTPA--PLPMPVAEDEIKPYIS 335  
  
QY 121 RCTVCEGPAIAVHSQTTDIPCPHGWISLWKGFSFIMFTSAGSGTGQAALASPGSCLE 180  
DB 336 RCSVCEAPAIATAVHSQDVSIHPCPAGWSSLWIGYFAMHTGAGAGGQGSLSVPSGSCLE 395  
  
QY 181 EFRASPFLECH-RGTCNNYSNSYFWLASLNPERMFR-KPIPTVKGABLEKIIISRCOV 238  
DB 396 DFRATPIECNGRGRTCHYANKYSFWLTTI-PEQSQGSPSADTLKAGLIRTHISRCOV 454  
  
QY 239 CMK 241  
DB 455 CMK 457

Mon Apr 5 07:53:04 2004

us-10-032-221b-10.rapb

Page 7

Search completed: April 5, 2004, 07:36:05  
Job time : 154.017 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 57.3075 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221b-10  
Perfect score: 1340  
Sequence: 1 GLKRGDGSPTWTRGPF.....KAGELEKLSRCQVCMKKRH 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1314	98.1	268	4	US-09-589-927-6
2	1314	98.1	268	4	US-09-277-665-6
3	1314	98.1	268	4	US-09-589-987-6
4	1210.5	90.3	471	2	US-08-399-889-24
5	1210.5	90.3	471	3	US-09-167-364-24
6	1210.5	90.3	471	3	US-09-439-897-2
7	1192	89.0	218	2	US-08-399-889-25
8	1192	89.0	218	3	US-09-167-364-25
9	1192	89.0	218	3	US-09-439-897-4
10	939	70.1	260	4	US-09-589-927-2
11	939	70.1	260	4	US-09-277-665-2
12	939	70.1	260	4	US-09-589-987-2
13	938.5	70.0	264	4	US-09-589-927-10
14	938.5	70.0	264	4	US-09-277-665-10
15	938.5	70.0	264	4	US-09-589-987-10
16	880	65.7	211	4	US-09-512-563C-46
17	741.5	55.3	258	4	US-09-589-927-4
18	741.5	55.3	258	4	US-09-277-665-4
19	741.5	55.3	258	4	US-09-589-987-4
20	727	54.3	260	4	US-09-589-927-12
21	727	54.3	260	4	US-09-277-665-12
22	727	54.3	260	4	US-09-589-987-12
23	690.5	51.5	260	4	US-09-589-927-8
24	690.5	51.5	260	4	US-09-277-665-8
25	690.5	51.5	260	4	US-09-589-987-8
26	339.5	25.3	1694	1	US-08-494-168-2
27	334	24.9	68	4	US-09-512-563C-50

28	334	24.9	72	4	US-09-512-563C-48	Sequence 48, Appl
29	334	24.9	72	4	US-09-512-563C-52	Sequence 52, Appl
30	328	24.5	72	4	US-09-512-563C-61	Sequence 61, Appl
31	189	14.1	36	3	US-09-439-897-65	Sequence 65, Appl
32	143	10.7	26	3	US-09-439-897-63	Sequence 63, Appl
33	116	8.7	21	4	US-09-512-563C-26	Sequence 26, Appl
34	113	8.4	21	4	US-09-512-563C-27	Sequence 27, Appl
35	110	8.2	35	3	US-09-439-897-64	Sequence 64, Appl
36	98	7.3	15	3	US-09-439-897-61	Sequence 61, Appl
37	85.5	6.4	409	4	US-09-198-452A-554	Sequence 554, App
38	83.5	6.2	404	1	US-07-602-824A-4	Sequence 4, Appl
39	83.5	6.2	404	1	US-07-602-808-4	Sequence 4, Appl
40	83.5	6.2	404	1	US-07-983-451-4	Sequence 4, Appl
41	83.5	6.2	404	1	US-08-261-578-4	Sequence 4, Appl
42	83.5	6.2	404	1	US-08-261-577-10	Sequence 10, Appl
43	83.5	6.2	539	6	5449756-4	Patent No. 5449756
44	82	6.1	15	3	US-09-439-897-53	Sequence 53, Appl
45	82	6.1	15	3	US-09-439-897-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1  
US-09-589-927-6  
; Sequence 6, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 945251  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 268  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-6

Query Match 98.1%; Score 1314; DB 4; Length 268;  
Best Local Similarity 99.6%; Pred. No. 8.4e-138;  
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	5	KRGDSGSPATWTTTRGFVTRHSQTALPSCPEGTVPLYSGFSLFVQGNORAHGQDLGTL	64
Db	29	KRGDSGSPATWTTTRGFVTRHSQTALPSCPEGTVPLYSGFSLFVQGNORAHGQDLGTL	88
QY	65	GSCLORFETMPFLFCNVNDVNCNFASRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTV	124
Db	89	GSCLORFETMPFLFCNVNDVNCNFASRNDYSYWLSTPALMPNMNAPITGRALEPYISRCTV	148
QY	125	CEGPAIAVAHSQTDIDPPCPHGWISLWKGFSFIMFTSAGSEGTQALAPGSCLEEFRA	184
Db	149	CEGPAIAVAHSQTDIDPPCPHGWISLWKGFSFIMFTSAGSEGTQALAPGSCLEEFRA	208
QY	185	SPFLECHGRGTCNYSNSYSYFWLASLNPENFRKPISTVKAGELEKIIISRCQVCMKKRH	244
Db	209	SPFLECHGRGTCNYSNSYSYFWLASLNPENFRKPISTVKAGELEKIIISRCQVCMKKRH	268

RESULT 2  
US-09-277-665-6  
; Sequence 6, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665



```
/ CURRENT FILING DATE: 1999-03-26
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 268
/ TYPE: PRT
/ ORGANISM: Human
US-09-277-663-6

Query Match      98.1%; Score 1314; DB 4; Length 268;
Best Local Similarity 99.6%; Pred. No. 8.4e-138;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGDSPATWTTTRGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQDLGTL 64
DB 29 KRGDSPATWTTTRGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQDLGTL 88

QY 65 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYISRCTV 148

QY 125 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCLEEFRA 184
DB 149 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCLEEFRA 208

QY 185 SPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVCKMKRH 244
DB 209 SPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVCKMKRH 268

RESULT 3
US-09-589-987-6
/ Sequence 6, Application US/0959987
/ Patent No. 6498140
/ GENERAL INFORMATION:
/ APPLICANT: University of Kansas Medical Center
/ TITLE OF INVENTION: The use of isolated domains of type IV collagen to
/ FILE REFERENCE: 945251
/ CURRENT FILING DATE: 2000-06-07
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 268
/ TYPE: PRT
/ ORGANISM: Human
US-09-589-987-6

Query Match      98.1%; Score 1314; DB 4; Length 268;
Best Local Similarity 99.6%; Pred. No. 8.4e-138;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRGDSPATWTTTRGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQDLGTL 64
DB 29 KRGDSPATWTTTRGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQDLGTL 88

QY 65 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYISRCTV 124
DB 89 GSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYISRCTV 148

QY 125 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCLEEFRA 184
DB 149 CEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCLEEFRA 208

QY 185 SPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVCKMKRH 244
DB 209 SPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVCKMKRH 268

RESULT 4
US-08-399-889-24
/ Sequence 24, Application US/08399889B

/ Patent No. 5973120
/ GENERAL INFORMATION:
/ APPLICANT: Reeder, Stephen T
/ APPLICANT: Morrison, Karen E
/ APPLICANT: Hudson, Billy G
/ TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
/ FILE REFERENCE: 951263A
/ CURRENT APPLICATION NUMBER: US/08/399,889B
/ CURRENT FILING DATE: 1995-03-07
/ EARLIER APPLICATION NUMBER: 07/621091
/ EARLIER FILING DATE: 1990-11-30
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 24
/ LENGTH: 471
/ TYPE: PRT
/ ORGANISM: Calf
US-08-399-889-24

Query Match      90.3%; Score 1210.5; DB 2; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRGDSGPATWTT-RGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQ 59
DB 227 GLKGRGDSGPATWTT-RGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQ 286

QY 60 DLGTGLGSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYI 119
DB 287 DLGTGLGSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNMWAPITGRALEPYI 346

QY 120 SRTVCCEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCL 179
DB 347 SRTVCCEGPAIAIAVHSQTTDIPCPCHGWISLWKGFIMFTSAGSEGTQALASPGSCL 406

QY 180 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVC 239
DB 407 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTSVKAGELEKIISRCQVC 466

QY 240 MKKR 243
DB 467 MKKR 470

RESULT 5
US-09-167-364-24
/ Sequence 24, Application US/09167364
/ Patent No. 6007980
/ GENERAL INFORMATION:
/ APPLICANT: Reeder, Stephen T
/ APPLICANT: Morrison, Karen E
/ APPLICANT: Hudson, Billy G
/ TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
/ FILE REFERENCE: 951263B
/ CURRENT APPLICATION NUMBER: US/09/167,364
/ CURRENT FILING DATE: 1998-10-07
/ EARLIER APPLICATION NUMBER: 08/399889
/ EARLIER FILING DATE: 1995-03-07
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 24
/ LENGTH: 471
/ TYPE: PRT
/ ORGANISM: Calf
US-09-167-364-24

Query Match      90.3%; Score 1210.5; DB 3; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRGDSGPATWTT-RGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQ 59
DB 227 GLKGRGDSGPATWTT-RGFVTRHSQTTPSCPEGTPLVSGSFLVQGNQRAHGQ 286
```



```
RESULT 9
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      89.0%; Score 1192; DB 3; Length 218;
Best Local Similarity 99.5%; Pred. No. 2.3e-124;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 QTTAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLOQRTTMTDFLCNVNDVCN 96
DB 1 QTTAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLOQRTTMTDFLCNVNDVCN 60

QY 87 FASRNDYSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

QY 147 GWSLWKGFIFNFTSAGSEGTQALASPGSCLEEFRAFPFLECHGRGTCNTYSNSYSFW 206
DB 121 GWSLWKGFIFNFTSAGSEGTQALASPGSCLEEFRAFPFLECHGRGTCNTYSNSYSFW 180

QY 207 LASLNPERMRKPIPTSTVKAGELEKIISRCQVCMKKH 244
DB 181 LASLNPERMRKPIPTSTVKAGELEKIISRCQVCMKKRH 218

RESULT 10
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match      70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLOQRTTMTDFLF 78
DB 36 GFLVTRHSQTTDIPPCPGTGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTSTMPFLF 95

QY 79 CNVNDVCNFAERNDSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQ 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPSMAPITGENTRPFISRCVCAEAPAMVAVHSQ 155

QY 139 TDIPCPHGWISLWKGFIFNFTSAGSEGTQALASPGSCLEEFRAFPFLECHGRGTCNY 198
DB 156 IQIPPCPGSWSSLIWGYFVMTSAGAGSQQALASPGSCLEEFRAFPFLECHGRGTCNY 215

RESULT 11
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

Query Match      70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLOQRTTMTDFLF 78
DB 36 GFLVTRHSQTTDIPPCPGTGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTSTMPFLF 95

QY 79 CNVNDVCNFAERNDSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQ 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPSMAPITGENTRPFISRCVCAEAPAMVAVHSQ 155

QY 139 TDIPCPHGWISLWKGFIFNFTSAGSEGTQALASPGSCLEEFRAFPFLECHGRGTCNY 198
DB 156 IQIPPCPGSWSSLIWGYFVMTSAGAGSQQALASPGSCLEEFRAFPFLECHGRGTCNY 215

RESULT 12
US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-2

Query Match      70.1%; Score 939; DB 4; Length 260;
Best Local Similarity 71.9%; Pred. No. 3.9e-96;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 19 GFVTRHSQTTAIPSCPGTVPVLYSGFSLVQGNQRAHQDGLTGLSCLOQRTTMTDFLF 78
DB 36 GFLVTRHSQTTDIPPCPGTGKILYHGYSLLYVQGNRAHQDGLTAGSCLRKFTSTMPFLF 95

QY 79 CNVNDVCNFAERNDSYWLSTPALPMNMAPITGRALEPIYISRCTVCEGPAIAIAVHSQ 138
DB 96 CNINNVNCFASRNDYSYWLSTPEPMPSMAPITGENTRPFISRCVCAEAPAMVAVHSQ 155

QY 139 TDIPCPHGWISLWKGFIFNFTSAGSEGTQALASPGSCLEEFRAFPFLECHGRGTCNY 198
DB 156 IQIPPCPGSWSSLIWGYFVMTSAGAGSQQALASPGSCLEEFRAFPFLECHGRGTCNY 215
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.14528 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221B-37

Perfect score: 147

Sequence: 1 TMPFLFCNVNDVCFNFSRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	100.0	220	2 B49736	collagen alpha 3(I
2	147	100.0	1670	1 CGHU3B	collagen alpha 3(I
3	146	99.3	471	2 A39024	collagen alpha 3(I
4	141	95.9	161	2 S49488	collagen alpha 3(I
5	141	95.9	246	2 I48302	collagen alpha 3(I
6	141	95.9	258	2 B41228	collagen alpha 1(I
7	141	95.9	1669	1 CGHU4B	collagen alpha 1(I
8	141	95.9	1669	1 CGMS4B	collagen alpha 1(I
9	139	94.6	253	2 I48304	collagen alpha 5(I
10	139	94.6	754	2 A5267	collagen alpha 5(I
11	139	94.6	1691	1 S22917	collagen alpha 5(I
12	129	87.8	1752	2 A45407	collagen alpha 3(I
13	125	85.0	1747	2 A4121	collagen alpha-4 c
14	125	85.0	1763	2 S46366	collagen alpha 2(I
15	120	81.6	1758	2 T29350	hypothetical prote
16	120	81.6	1759	2 T29351	collagen alpha 2(I
17	117	79.6	1744	2 S40991	collagen alpha 1(I
18	116	78.9	261	2 A34476	collagen alpha 2(I
19	115	78.2	1691	1 CGHU6B	collagen alpha 6(I
20	112	76.2	1775	2 A41228	collagen alpha 2(I
21	112	76.2	1707	2 A33526	collagen alpha 2(I
22	112	76.2	1712	1 CGHU2B	collagen alpha 4(I
23	103	70.1	312	2 I48303	collagen alpha 4(I
24	103	70.1	623	2 A45137	collagen alpha 4(I
25	103	70.1	1690	1 CGHU1B	collagen alpha 4(I
26	102	69.4	453	2 S48804	collagen alpha 4(I
27	93	63.3	1775	2 A18993	collagen alpha 1(I
28	87	59.2	1761	2 T13990	collagen type IV a
29	73	49.7	79	2 C43928	probable collagen

#### ALIGNMENTS

##### RESULT 1

B49736  
collagen alpha 3(IV) chain, medium splice form - human (fragment)  
N;Contains: collagen alpha 3(IV) chain, splice form GP-V  
C;Species: Homo sapiens (man)  
C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text\_change 17-Mar-2000  
C;Accession: B49736; D49736; S69111  
R;Peng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.  
A;Reference number: A49736; MUID:94124597; PMID:8294492  
A;Accession: B49736  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 169-220 <FEN1>  
A;Accession: D49736  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: mRNA  
A;Residues: 22-220 <FEN2>  
A;Cross-references: GB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107  
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank  
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wi  
Eur. J. Biochem. 229, 754-760, 1995  
A;Title: Characterization and expression of multiple alternatively spliced transcripts  
A;Accession: S69111; MUID:95278230; PMID:7758473  
A;Reference number: S69111  
A;Accession: S69111  
A;Molecule type: mRNA  
A;Residues: 1-45,169-204,'L',206-220 <PEN>  
A;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
A;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace  
F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pred  
F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status  
F;22-220/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 100.0%; Score 147; DB 2; Length 220;

Best Local Similarity 100.0%; Pred. No. 1.8e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCFNFSRNDYSYWL 25

DB 82 TMPFLFCNVNDVCFNFSRNDYSYWL 106

##### RESULT 2

CGHU3B

F;43-1438/Region: interrupted helical  
F;791-793/Region: cell attachment (R-G-D) motif  
F;996-998/Region: cell attachment (R-G-D) motif  
F;1154-1156/Region: cell attachment (R-G-D) motif  
F;1306-1308/Region: cell attachment (R-G-D) motif  
F;1345-1347/Region: cell attachment (R-G-D) motif  
F;1432-1434/Region: cell attachment (R-G-D) motif  
F;1439-1670/Domain: carboxyl-terminal nonhelical NCI <NC1>  
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTL>  
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CR2>  
F;1631-33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status pred  
F;1460-1548,1493-1551/disulfide bonds: (covalent) #status predicted  
F;1505-1511,1616-1622/disulfide bonds: #status predicted  
F;1570-1622,1604-1665/disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 147; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 1.2e-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPLFCNVNDVCFASRNDYSYLW 25  
| | | | | | | | | | | | | | | | | | | |  
Db 1499 TMPLFCNVNDVCFASRNDYSYLW 1523

RESULT 3  
A39024  
collagen alpha 3(IV) chain - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text change 13-Aug-1999  
C;Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815  
R;Morrison, K.E.; Garmino, G.G.; Reeders, S.T.  
J. Biol. Chem. 266, 34-39, 1991  
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
A;Reference number: A39024; MUID:91093146; PMID:1985905  
A;Accession: A39024  
A;Molecule type: mRNA  
A;Residues: 1-471 <WOR>  
A;Cross-references: EMBL:M63139; NID:G162886; PIDN:AAA62708.1; PID:g162887  
R;Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
A;Reference number: S18432; MUID:87222419; PMID:2438283  
A;Accession: S20672  
A;Molecule type: protein  
A;Residues: 227-228,'X',230-244 <BUT>  
R;Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
J. Biol. Chem. 263, 13374-13380, 1988  
A;Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
A;Reference number: S17802; MUID:88330844; PMID:3417661  
A;Accession: S17802  
A;Molecule type: protein  
A;Residues: 227-228,'X',230-252,'Y',254 <SAU>  
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5466-5469, 1990  
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of typ  
A;Reference number: A35167; MUID:90202779; PMID:2318822  
A;Accession: A35167  
A;Molecule type: protein  
A;Residues: 236-258 <GUN>  
R;Gunwar, S.; Bailester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No  
J. Biol. Chem. 266, 15318-15324, 1991  
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol  
A;Reference number: A39419; MUID:91332055; PMID:1869555  
A;Accession: C39419  
A;Molecule type: protein  
A;Residues: 236-255 <GU2>  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;  
F;1-238/Domain: collagenous (fragment) #status predicted <COI>  
F;239-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>  
F;239-353/Domain: repeat NC1 #status predicted <NC1>  
F;354-471/Domain: repeat NCI #status predicted <NC12>

A; Note: the amino end of the mature form is blocked  
R/Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1987



A:Title: Complete primary structure of the alpha-1(I)-chain of human basement membrane (ty  
A:Reference number: S00207; MUID:88083584; PMID:3691802  
A:Accession: S00207  
A:Molecule type: mRNA  
A:Residues: 244-530 <S013>  
A:Cross-references: EMBL:Y00706; NID:g29548; PIDN:CAA68698.1; PID:g29549  
R:Esle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A:Title: The alpha-1(I)-chain of human basement membrane (type IV) collagen. Complete amino-acid se  
A:Reference number: S39614; MUID:94038963; PMID:8223498  
A:Accession: S39614  
A:Molecule type: protein  
A:Residues: 371-554 <EBL>  
R:Babel, W.; Glanville, R.W.  
Eur. J. Biochem. 143, 545-556, 1984  
A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se  
A:Reference number: A02863; MUID:85003629; PMID:6434307  
A:Accession: A02863  
A:Molecule type: protein  
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interr  
A:Reference number: S16908; MUID:82005835; PMID:6792033  
A:Accession: A58517  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14  
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (b  
A:Reference number: S16910; MUID:84053346; PMID:6416291  
A:Accession: S16910  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-  
R:Phillips, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.  
A:Reference number: S16879; MUID:85216555; PMID:2582422  
A:Accession: S16879  
A:Molecule type: mRNA  
A:Residues: 1259-1669 <BRI>  
A:Cross-references: EMBL:M10940; NID:g180421; PIDN:AAA52006.1; PID:g180424  
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm  
A:Reference number: S02550; MUID:89005112; PMID:2844531  
A:Contents: annotation; disulfide bonds  
C:Genetics:  
A:Gene: GDB:COL4A1  
A:Cross-references: GDB:119791; OMIM:120130  
A:Map position: 13q34-13q34  
A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/  
A:Exons: 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116  
A:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(I  
A:Associations: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim  
A:Associations: in the interrupted helical domain (with disulfide and desmosine cr

C:Function:  
A:Description: structural component of extracellular basement membrane  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication;  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F:29-162/Domain: amino-terminal nonhelical, 78 <7SD>  
F:163-1440/Domain: interrupted helical <COL>  
F:414-452/Region: integrin binding #status experimental  
F:597-599/Region: cell attachment (R-G-D) motif  
F:917-919/Region: cell attachment (R-G-D) motif  
F:968-970/Region: cell attachment (R-G-D) motif  
F:1441-1669/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTI>  
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTI>  
F:27/Modified site: blocked amino end (Aia) (in mature form) #status experimental  
F:31.36.39, 41, 125, 434, 467, 470/Disulfide bonds: interchain #status predicted  
F:45.48, 78, 90, 129, 156, 172, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 540, 543, 5  
1081, 1084, 1099, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 13  
F:45.48, 78, 90, 129, 156, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 6  
99, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 1340, 1356, 1371  
F:54.62, 75, 84, 87, 96, 102, 105, 108, 111, 117, 120, 123, 135, 141, 147, 150, 153, 159, 167, 178, 181, 184  
F:419, 422, 425, 439, 445, 448, 451, 479, 485, 491, 494, 503, 512, 518, 524, 530, 546, 549, 552, 555, 561, 56  
9, 745, 748, 751, 754, 763/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:126/Binding site: carboxylate (Asn) (covalent) #status experimental  
F:129/Modified site: allysine (Lys) #status predicted  
F:172.540, 947/Modified site: 5-hydroxylysine (Lys) #status atypical  
F:272.645, 839/Modified site: 4-hydroxyproline (Pro) #status atypical  
F:466.775, 784, 787, 799, 796, 804, 810, 816, 822, 834, 860, 863, 869, 872, 875, 887, 890, 893, 899, 9  
F:31.129, 1138, 1141, 1159, 1171, 1176, 1179, 1194, 1200, 1203, 1215, 1224, 1227, 1244, 1247, 1250, 1256  
431, 1437/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:1120, 1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
F:1120, 1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental  
F:1215, 1224/Modified site: 3-hydroxyproline (Pro) #status absent  
F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F:1505-1511, 1616-1622/Disulfide bonds: #status predicted  
F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
Query Match 95.9%; Score 141; DB 1; Length 1669;  
Best Local Similarity 92.0%; Pred No. 8.3e-12;  
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TMPPFLFCNVNVCNFCASNDYSYWL 25  
Db 1499 TMPPFLFCNVNVCNFCASNDYSYWL 1523  
RESULT 8  
CGMS4B  
collagen alpha 1(IV) chain precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000  
C:Accession: A03525; S01454; A28066; A25636; A29301; S19079; A32003; A31766; S1  
R:Mutukumar, G.; Blumberg, B.; Kurkinen, M.  
J. Biol. Chem. 264, 6310-6317, 1989  
A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Dif  
A:Reference number: A33525; MUID:89197932; PMID:2703490  
A:Accession: A33525  
A:Molecule type: mRNA  
A:Residues: 1-1669 <MUT>  
A:Cross-references: EMBL:J04694; NID:g556296; PIDN:AAA50292.1; PID:g556297  
R:Wood, L.; Thériault, N.; Vogeli, G.  
FEBS Lett. 227, 5-8, 1988  
A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the a  
A:Reference number: S01454; MUID:88112221; PMID:3338568  
A:Accession: S01454  
A:Molecule type: mRNA  
A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-7  
A:Cross-references: EMBL:X06777  
R:Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.  
J. Biol. Chem. 263, 8706-8709, 1988

A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
A;Reference number: A28066; MUID:88243724; PMID:3379041  
A;Accession: A28066  
A;Molecule type: mRNA  
A;Residues: 1-129 <K11>  
A;Cross-references: EMBL:J03758; NID:g192669; PIDN:AAA37439.1; PID:g192670  
R;Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1995  
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1(IV) collagen chain  
A;Reference number: A02864; MUID:85127033; PMID:2578961  
A;Accession: A02864  
A;Molecule type: mRNA  
A;Residues: 1276-1669 <OBE>  
A;Cross-references: EMBL:X02201; NID:g50233; PIDN:CAA26132.1; PID:g1333876  
R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G. Gene 43, 301-304, 1986  
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox  
A;Reference number: A25636; MUID:86301886; PMID:3755692  
A;Accession: A25636  
A;Molecule type: mRNA  
A;Residues: 1149-1396, 'S', 1398-1424 <NAT>  
A;Cross-references: EMBL:M14042; NID:g192286; PIDN:AAA37342.1; PID:g192287  
A;Note: The authors translated the codon CAG for residue 1374 as ARG  
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj J. Biol. Chem. 262, 8496-8499, 1987  
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV) collagen and human alpha-1(IV) collagen  
A;Reference number: A94680; MUID:87250460; PMID:35978383  
A;Accession: A29301  
A;Molecule type: mRNA  
A;Residues: 1441-1669 <KUR>  
A;Cross-references: EMBL:M15832; NID:g192282; PIDN:AAA37340.1; PID:g387115  
R;Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y. J. Biol. Chem. 263, 12310-12314, 1988  
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequen  
A;Reference number: S19079; MUID:88315019; PMID:2842328  
A;Accession: S19079  
A;Molecule type: DNA  
A;Residues: 1-28 <K12>  
A;Cross-references: EMBL:J03944; NID:g192673; PIDN:AAA37442.1; PID:g466503  
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G. J. Biol. Chem. 263, 19274-19277, 1988  
A;Title: Head-to-head arrangement of murine type IV collagen genes.  
A;Reference number: A92702; MUID:89066738; PMID:3198626  
A;Accession: A32003  
A;Molecule type: DNA  
A;Residues: 1-28 <KAY>  
A;Cross-references: EMBL:J04448; NID:g192666; PIDN:AAA37437.1; PID:g450449  
R;Burbelo, P.D.; Martin, G.R.; Yamada, Y. Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
A;Reference number: A94220; MUID:89071759; PMID:3200851  
A;Accession: A31766  
A;Molecule type: DNA  
A;Residues: 1-28 <BUR>  
A;Cross-references: EMBL:M23333; NID:g340878; PIDN:AAA51625.1; PID:g535668  
R;Sakurai, Y.; Sullivan, M.; Yamada, Y. J. Biol. Chem. 261, 6654-6657, 1986  
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
A;Reference number: S19094; MUID:86196099; PMID:3009458  
A;Accession: S19094  
A;Molecule type: DNA  
A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>  
R;Schuppan, D.; Timpl, R.; Glanville, R.W. FEBS Lett. 115, 297-300, 1980  
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty  
A;Reference number: S16909; MUID:80246483; PMID:6772473  
A;Accession: S16909  
A;Molecule type: protein  
A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-1238  
R;Schuppan, D.; Glanville, R.W.; Timpl, R. Eur. J. Biochem. 123, 505-512, 1982  
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amir

A;Reference number: A25991; MUID:82186723; PMID:6804236  
A;Accession: A25991  
A;Molecule type: Protein  
A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 10  
61, 'X', 1063-1069, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119  
A;Accession: B25991  
A;Molecule type: Protein  
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234  
3, 'SP', 1266, 'IN', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 132  
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R. Eur. J. Biochem. 139, 401-410, 1984  
A;Title: Subunit structure and assembly of the globular domain of basement-membrane col  
A;Reference number: S17801; MUID:84132058; PMID:6698021  
A;Accession: S17801  
A;Molecule type: Protein  
A;Residues: 1435-1443 <WEB>  
C;Genetics:  
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3  
A;Note: The list of introns may be incomplete  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular ma  
F;1-27/DNA: signal sequence #status predicted <SIG>  
F;28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F;28-162/DNA: 'S' <7SD>  
F;163-1440/DNA: collagenous, triple helix <COL>  
F;597-599/Region: cell attachment (R-G-D) motif  
F;781-783/Region: cell attachment (R-G-D) motif  
F;917-919/Region: cell attachment (R-G-D) motif  
F;968-970/Region: cell attachment (R-G-D) motif  
F;1441-1669/DNA: carboxyl-terminal nonhelical, NCI <NCI>  
F;1441-1552/Region: duplication  
F;1553-1669/Region: duplication  
F;3136-39, 41, 434, 467, 470/Disulfide bonds: Interchain #status predicted  
F;126/Binding site: carboxylate (Asn) (covalent) #status predicted  
F;971-974, 977, 986, 989, 1001, 1007, 1019, 1022, 1031, 1037, 1040, 1055, 1060, 1063, 1075, 1078, 1090,  
92, 1298, 1310, 1313, 1322, 1337, 1346, 1349, 1422, 1425, 1431, 1437, 1440/Modified site: hydroxypro  
F;1214, 1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental  
F;1505-1511, 1616-1622/Disulfide bonds: #status predicted  
Query Match 95.9%; Score 141; DB 1; Length 1669;  
Best Local Similarity 92.0%; Pred. No. 8.3e-12;  
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TNPFLFCNVNVCNPFASNDYSYWL 25  
DB 1499 TNPFLFCNVNVCNPFASNDYSYWL 1523  
RESULT 9  
I48304  
collagen alpha 5(IV) chain - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999  
C;Accession: I48304; S47280  
R;Miner, J.H.; Sares, J.R. J. Cell Biol. 127, 879-891, 1994  
A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
A;Reference number: A54979; MUID:95050957; PMID:7962065  
A;Accession: I48304  
A;Molecule type: mRNA  
A;Residues: 1-253 <RES>  
A;Cross-references: EMBL:Z35168; NID:g535201; PIDN:CAA84531.1; PID:g535202  
C;Superfamily: collagen alpha 1(IV) chain  
Query Match 94.6%; Score 139; DB 2; Length 253;  
Best Local Similarity 88.0%; Pred. No. 2.8e-12;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TNPFLFCNVNVCNPFASNDYSYWL 25  
DB 83 TNPFLFCNVNVCNPFASNDYSYWL 107

Hum. Mol. Genet. 1, 127-129, 1992

A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in  
A:Reference number: I54317; MUID:93244772; PMID:1363780

A:Accession: I54317

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 313-324, 'E', 326-330 <REN>

A:Cross-references: GS:59334; NID:G939946; PIDN:AAD13909.1; PID:G4261509

R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeythvae, M.; Shows, T.B.; Tryggvason, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990

A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidn

A:Reference number: A34850; MUID:90160375; PMID:1689491

A:Accession: A34850

A:Molecule type: mRNA

A:Residues: 914-1264, 1271-1691 <HOS>

A:Cross-references: EMBL:M31115; NID:G180824; PIDN:AAA52045.1; PID:G180825

R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
Genomics 9, 1-9, 1991

A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene tha

A:Reference number: A37969; MUID:91169491; PMID:2004755

A:Accession: S18850

A:Molecule type: DNA

A:Residues: 924-1264, 1271-1691 <ZH3>

A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; E

8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:G177922; PIDN:AAA51558.1; P

R:Guo, C.B.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; I

Kidney Int. 44, 1316-1321, 1993

A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a comple

A:Reference number: I56971; MUID:94133540; PMID:8301933

A:Accession: I56971

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1258-1276 <GUOI>

A:Cross-references: GB:S69168; NID:G5450395; PIDN:AAC60612.1; PID:G5450396

A:Note: kidney splice form

A:Accession: I76598

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1284-1291, 'TFLGYACLV', <GUO2>

A:Cross-references: GB:S69169; NID:G545087; PIDN:AAC60613.1; PID:G545098

A:Note: frameshift mutation in patient with Alport syndrome

R:Myers, J.C.; Jones, T.A.; Pohjolaainen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; S

Am. J. Hum. Genet. 46, 1024-1033, 1990

A:Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the reg

A:Reference number: A35335; MUID:90252791; PMID:2339699

A:Accession: A35335

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1448-1477 <MYE>

R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yo

Kidney Int. 46, 1307-1314, 1994

A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primor

A:Reference number: I56975; MUID:95156893; PMID:7853788

A:Accession: I56975

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1595-1602 <NAK>

A:Cross-references: GB:S75903; NID:G913882; PIDN:AAB33374.1; PID:G913883

A:Note: peracute termination mutation from a patient with Alport syndrome; one other m

R:Lemnick, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;

Genomics 17, 485-489, 1993

A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Alp

A:Reference number: I54188; MUID:94010948; PMID:8406498

A:Accession: I54188

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1604-1607, 'VHDAYKC', <LEM>

A:Cross-references: GB:S65767; NID:G425563; PIDN:AAD13967.1; PID:G4261567

A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations a

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

C:Genetics:

A:Gene: GDB:COL4A5; ATS

A;Cross-references: GDB:120596; OMIM:303630  
A;Map position: X222-Xq22  
A;Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1191/1; 1217/1; 1244/1; 1271/1; 1298/1; 1325/1; 1352/1; 1379/1; 1406/1; 1433/1; 1460/1; 1487/1; 1514/1; 1541/1; 1568/1; 1595/1; 1622/1; 1649/1; 1676/1; 1703/1; 1730/1; 1757/1; 1784/1; 1811/1; 1838/1; 1865/1; 1892/1; 1919/1; 1946/1; 1973/1; 2000/1; 2027/1; 2054/1; 2081/1; 2108/1; 2135/1; 2162/1; 2189/1; 2216/1; 2243/1; 2270/1; 2297/1; 2324/1; 2351/1; 2378/1; 2405/1; 2432/1; 2459/1; 2486/1; 2513/1; 2540/1; 2567/1; 2594/1; 2621/1; 2648/1; 2675/1; 2702/1; 2729/1; 2756/1; 2783/1; 2810/1; 2837/1; 2864/1; 2891/1; 2918/1; 2945/1; 2972/1; 2999/1; 3026/1; 3053/1; 3080/1; 3107/1; 3134/1; 3161/1; 3188/1; 3215/1; 3242/1; 3269/1; 3296/1; 3323/1; 3350/1; 3377/1; 3404/1; 3431/1; 3458/1; 3485/1; 3512/1; 3539/1; 3566/1; 3593/1; 3620/1; 3647/1; 3674/1; 3701/1; 3728/1; 3755/1; 3782/1; 3809/1; 3836/1; 3863/1; 3890/1; 3917/1; 3944/1; 3971/1; 3998/1; 4025/1; 4052/1; 4079/1; 4106/1; 4133/1; 4160/1; 4187/1; 4214/1; 4241/1; 4268/1; 4295/1; 4322/1; 4349/1; 4376/1; 4403/1; 4430/1; 4457/1; 4484/1; 4511/1; 4538/1; 4565/1; 4592/1; 4619/1; 4646/1; 4673/1; 4700/1; 4727/1; 4754/1; 4781/1; 4808/1; 4835/1; 4862/1; 4889/1; 4916/1; 4943/1; 4970/1; 4997/1; 5024/1; 5051/1; 5078/1; 5105/1; 5132/1; 5159/1; 5186/1; 5213/1; 5240/1; 5267/1; 5294/1; 5321/1; 5348/1; 5375/1; 5402/1; 5429/1; 5456/1; 5483/1; 5510/1; 5537/1; 5564/1; 5591/1; 5618/1; 5645/1; 5672/1; 5699/1; 5726/1; 5753/1; 5780/1; 5807/1; 5834/1; 5861/1; 5888/1; 5915/1; 5942/1; 5969/1; 5996/1; 6023/1; 6050/1; 6077/1; 6104/1; 6131/1; 6158/1; 6185/1; 6212/1; 6239/1; 6266/1; 6293/1; 6320/1; 6347/1; 6374/1; 6401/1; 6428/1; 6455/1; 6482/1; 6509/1; 6536/1; 6563/1; 6590/1; 6617/1; 6644/1; 6671/1; 6698/1; 6725/1; 6752/1; 6779/1; 6806/1; 6833/1; 6860/1; 6887/1; 6914/1; 6941/1; 6968/1; 6995/1; 7022/1; 7049/1; 7076/1; 7103/1; 7130/1; 7157/1; 7184/1; 7211/1; 7238/1; 7265/1; 7292/1; 7319/1; 7346/1; 7373/1; 7400/1; 7427/1; 7454/1; 7481/1; 7508/1; 7535/1; 7562/1; 7589/1; 7616/1; 7643/1; 7670/1; 7697/1; 7724/1; 7751/1; 7778/1; 7805/1; 7832/1; 7859/1; 7886/1; 7913/1; 7940/1; 7967/1; 7994/1; 8021/1; 8048/1; 8075/1; 8102/1; 8129/1; 8156/1; 8183/1; 8210/1; 8237/1; 8264/1; 8291/1; 8318/1; 8345/1; 8372/1; 8399/1; 8426/1; 8453/1; 8480/1; 8507/1; 8534/1; 8561/1; 8588/1; 8615/1; 8642/1; 8669/1; 8696/1; 8723/1; 8750/1; 8777/1; 8804/1; 8831/1; 8858/1; 8885/1; 8912/1; 8939/1; 8966/1; 8993/1; 9020/1; 9047/1; 9074/1; 9101/1; 9128/1; 9155/1; 9182/1; 9209/1; 9236/1; 9263/1; 9290/1; 9317/1; 9344/1; 9371/1; 9398/1; 9425/1; 9452/1; 9479/1; 9506/1; 9533/1; 9560/1; 9587/1; 9614/1; 9641/1; 9668/1; 9695/1; 9722/1; 9749/1; 9776/1; 9803/1; 9830/1; 9857/1; 9884/1; 9911/1; 9938/1; 9965/1; 9992/1; 10019/1; 10046/1; 10073/1; 10100/1; 10127/1; 10154/1; 10181/1; 10208/1; 10235/1; 10262/1; 10289/1; 10316/1; 10343/1; 10370/1; 10397/1; 10424/1; 10451/1; 10478/1; 10505/1; 10532/1; 10559/1; 10586/1; 10613/1; 10640/1; 10667/1; 10694/1; 10721/1; 10748/1; 10775/1; 10802/1; 10829/1; 10856/1; 10883/1; 10910/1; 10937/1; 10964/1; 10991/1; 11018/1; 11045/1; 11072/1; 11099/1; 11126/1; 11153/1; 11180/1; 11207/1; 11234/1; 11261/1; 11288/1; 11315/1; 11342/1; 11369/1; 11396/1; 11423/1; 11450/1; 11477/1; 11504/1; 11531/1; 11558/1; 11585/1; 11612/1; 11639/1; 11666/1; 11693/1; 11720/1; 11747/1; 11774/1; 11801/1; 11828/1; 11855/1; 11882/1; 11909/1; 11936/1; 11963/1; 11990/1; 12017/1; 12044/1; 12071/1; 12098/1; 12125/1; 12152/1; 12179/1; 12206/1; 12233/1; 12260/1; 12287/1; 12314/1; 12341/1; 12368/1; 12395/1; 12422/1; 12449/1; 12476/1; 12503/1; 12530/1; 12557/1; 12584/1; 12611/1; 12638/1; 12665/1; 12692/1; 12719/1; 12746/1; 12773/1; 12800/1; 12827/1; 12854/1; 12881/1; 12908/1; 12935/1; 12962/1; 12989/1; 13016/1; 13043/1; 13070/1; 13097/1; 13124/1; 13151/1; 13178/1; 13205/1; 13232/1; 13259/1; 13286/1; 13313/1; 13340/1; 13367/1; 13394/1; 13421/1; 13448/1; 13475/1; 13502/1; 13529/1; 13556/1; 13583/1; 13610/1; 13637/1; 13664/1; 13691/1; 13718/1; 13745/1; 13772/1; 13799/1; 13826/1; 13853/1; 13880/1; 13907/1; 13934/1; 13961/1; 13988/1; 14015/1; 14042/1; 14069/1; 14096/1; 14123/1; 14150/1; 14177/1; 14204/1; 14231/1; 14258/1; 14285/1; 14312/1; 14339/1; 14366/1; 14393/1; 14420/1; 14447/1; 14474/1; 14501/1; 14528/1; 14555/1; 14582/1; 14609/1; 14636/1; 14663/1; 14690/1; 14717/1; 14744/1; 14771/1; 14798/1; 14825/1; 14852/1; 14879/1; 14906/1; 14933/1; 14960/1; 14987/1; 15014/1; 15041/1; 15068/1; 15095/1; 15122/1; 15149/1; 15176/1; 15203/1; 15230/1; 15257/1; 15284/1; 15311/1; 15338/1; 15365/1; 15392/1; 15419/1; 15446/1; 15473/1; 15500/1; 15527/1; 15554/1; 15581/1; 15608/1; 15635/1; 15662/1; 15689/1; 15716/1; 15743/1; 15770/1; 15797/1; 15824/1; 15851/1; 15878/1; 15905/1; 15932/1; 15959/1; 15986/1; 16013/1; 16040/1; 16067/1; 16094/1; 16121/1; 16148/1; 16175/1; 16202/1; 16229/1; 16256/1; 16283/1; 16310/1; 16337/1; 16364/1; 16391/1; 16418/1; 16445/1; 16472/1; 16499/1; 16526/1; 16553/1; 16580/1; 16607/1; 16634/1; 16661/1; 16688/1; 16715/1; 16742/1; 16769/1; 16796/1; 16823/1; 16850/1; 16877/1; 16904/1; 16931/1; 16958/1; 16985/1; 17012/1; 17039/1; 17066/1; 17093/1; 17120/1; 17147/1; 17174/1; 17201/1; 17228/1; 17255/1; 17282/1; 17309/1; 17336/1; 17363/1; 17390/1; 17417/1; 17444/1; 17471/1; 17498/1; 17525/1; 17552/1; 17579/1; 17606/1; 17633/1; 17660/1; 17687/1; 17714/1; 17741/1; 17768/1; 17795/1; 17822/1; 17849/1; 17876/1; 17903/1; 17930/1; 17957/1; 17984/1; 18011/1; 18038/1; 18065/1; 18092/1; 18119/1; 18146/1; 18173/1; 18200/1; 18227/1; 18254/1; 18281/1; 18308/1; 18335/1; 18362/1; 18389/1; 18416/1; 18443/1; 18470/1; 18497/1; 18524/1; 18551/1; 18578/1; 18605/1; 18632/1; 18659/1; 18686/1; 18713/1; 18740/1; 18767/1; 18794/1; 18821/1; 18848/1; 18875/1; 18902/1; 18929/1; 18956/1; 18983/1; 19010/1; 19037/1; 19064/1; 19091/1; 19118/1; 19145/1; 19172/1; 19199/1; 19226/1; 19253/1; 19280/1; 19307/1; 19334/1; 19361/1; 19388/1; 19415/1; 19442/1; 19469/1; 19496/1; 19523/1; 19550/1; 19577/1; 19604/1; 19631/1; 19658/1; 19685/1; 19712/1; 19739/1; 19766/1; 19793/1; 19820/1; 19847/1; 19874/1; 19901/1; 19928/1; 19955/1; 19982/1; 20009/1; 20036/1; 20063/1; 20090/1; 20117/1; 20144/1; 20171/1; 20198/1; 20225/1; 20252/1; 20279/1; 20306/1; 20333/1; 20360/1; 20387/1; 20414/1; 20441/1; 20468/1; 20495/1; 20522/1; 20549/1; 20576/1; 20603/1; 20630/1; 20657/1; 20684/1; 20711/1; 20738/1; 20765/1; 20792/1; 20819/1; 20846/1; 20873/1; 20900/1; 20927/1; 20954/1; 20981/1; 21008/1; 21035/1; 21062/1; 21089/1; 21116/1; 21143/1; 21170/1; 21197/1; 21224/1; 21251/1; 21278/1; 21305/1; 21332/1; 21359/1; 21386/1; 21413/1; 21440/1; 21467/1; 21494/1; 21521/1; 21548/1; 21575/1; 21602/1; 21629/1; 21656/1; 21683/1; 21710/1; 21737/1; 21764/1; 21791/1; 21818/1; 21845/1; 21872/1; 21899/1; 21926/1; 21953/1; 21980/1; 22007/1; 22034/1; 22061/1; 22088/1; 22115/1; 22142/1; 22169/1; 22196/1; 22223/1; 22250/1; 22277/1; 22304/1; 22331/1; 22358/1; 22385/1; 22412/1; 22439/1; 22466/1; 22493/1; 22520/1; 22547/1; 22574/1; 22601/1; 22628/1; 22655/1; 22682/1; 22709/1; 22736/1; 22763/1; 22790/1; 22817/1; 22844/1; 22871/1; 22898/1; 22925/1; 22952/1; 22979/1; 23006/1; 23033/1; 23060/1; 23087/1; 23114/1; 23141/1; 23168/1; 23195/1; 23222/1; 23249/1; 23276/1; 23303/1; 23330/1; 23357/1; 23384/1; 23411/1; 23438/1; 23465/1; 23492/1; 23519/1; 23546/1; 23573/1; 23600/1; 23627/1; 23654/1; 23681/1; 23708/1; 23735/1; 23762/1; 23789/1; 23816/1; 23843/1; 23870/1; 23897/1; 23924/1; 23951/1; 23978/1; 24005/1; 24032/1; 24059/1; 24086/1; 24113/1; 24140/1; 24167/1; 24194/1; 24221/1; 24248/1; 24275/1; 24302/1; 24329/1; 24356/1; 24383/1; 24410/1; 24437/1; 24464/1; 24491/1; 24518/1; 24545/1; 24572/1; 24599/1; 24626/1; 24653/1; 24680/1; 24707/1; 24734/1; 24761/1; 24788/1; 24815/1; 24842/1; 24869/1; 24896/1; 24923/1; 24950/1; 24977/1; 25004/1; 25031/1; 25058/1; 25085/1; 25112/1; 25139/1; 25166/1; 25193/1; 25220/1; 25247/1; 25274/1; 25301/1; 25328/1; 25355/1; 25382/1; 25409/1; 25436/1; 25463/1; 25490/1; 25517/1; 25544/1; 25571/1; 25598/1; 25625/1; 25652/1; 25679/1; 25706/1; 25733/1; 25760/1; 25787/1; 25814/1; 25841/1; 25868/1; 25895/1; 25922/1; 25949/1; 25976/1; 26003/1; 26030/1; 26057/1; 26084/1; 26111/1; 26138/1; 26165/1; 26192/1; 26219/1; 26246/1; 26273/1; 26300/1; 26327/1; 26354/1; 26381/1; 26408/1; 26435/1; 26462/1; 26489/1; 26516/1; 26543/1; 26570/1; 26597/1; 26624/1; 26651/1; 26678/1; 26705/1; 26732/1; 26759/1; 26786/1; 26813/1; 26840/1; 26867/1; 26894/1; 26921/1; 26948/1; 26975/1; 27002/1; 27029/1; 27056/1; 27083/1; 27110/1; 27137/1; 27164/1; 27191/1; 27218/1; 27245/1; 27272/1; 27299/1; 27326/1; 27353/1; 27380/1; 27407/1; 27434/1; 27461/1; 27488/1; 27515/1; 27542/1; 27569/1; 27596/1; 27623/1; 27650/1; 27677/1; 27704/1; 27731/1; 27758/1; 27785/1; 27812/1; 27839/1; 27866/1; 27893/1; 27920/1; 27947/1; 27974/1; 28001/1; 28028/1; 28055/1; 28082/1; 28109/1; 28136/1; 28163/1; 28190/1; 28217/1; 28244/1; 28271/1; 28298/1; 28325/1; 28352/1; 28379/1; 28406/1; 28433/1; 28460/1; 28487/1; 28514/1; 28541/1; 28568/1; 28595/1; 28622/1; 28649/1; 28676/1; 28703/1; 28730/1; 28757/1; 28784/1; 28811/1; 28838/1; 28865/1; 28892/1; 28919/1; 28946/1; 28973/1; 29000/1; 29027/1; 29054/1; 29081/1; 29108/1; 29135/1; 29162/1; 29189/1; 29216/1; 29243/1; 29270/1; 29297/1; 29324/1; 29351/1; 29378/1; 29405/1; 29432/1; 29459/1; 29486/1; 29513/1; 29540/1; 29567/1; 29594/1; 29621/1; 29648/1; 29675/1; 29702/1; 29729/1; 29756/1; 29783/1; 29810/1; 29837/1; 29864/1; 29891/1; 29918/1; 29945/1; 29972/1; 30000/1; 30027/1; 30054/1; 30081/1; 30108/1; 30135/1; 30162/1; 30189/1; 30216/1; 30243/1; 30270/1; 30297/1; 30324/1; 30351/1; 30378/1; 30405/1; 30432/1; 30459/1; 30486/1; 30513/1; 30540/1; 30567/1; 30594/1; 30621/1; 30648/1; 30675/1; 30702/1; 30729/1; 30756/1; 30783/1; 30810/1; 30837/1; 30864/1; 30891/1; 30918/1; 30945/1; 30972/1; 31000/1; 31027/1; 31054/1; 31081/1; 31108/1; 31135/1; 31162/1; 31189/1; 31216/1; 31243/1; 31270/1; 31297/1; 31324/1; 31351/1; 31378/1; 31405/1; 31432/1; 31459/1; 31486/1; 31513/1; 31540/1; 31567/1; 31594/1; 31621/1; 31648/1; 31675/1; 31702/1; 31729/1; 31756/1; 31783/1; 31810/1; 31837/1; 31864/1; 31891/1; 31918/1; 31945/1; 31972/1; 32000/1; 32027/1; 32054/1; 32081/1; 32108/1; 32135/1; 32162/1; 32189/1; 32216/1; 32243/1; 32270/1; 32297/1; 32324/1; 32351/1; 32378/1; 32405/1; 32432/1; 32459/1; 32486/1; 32513/1; 32540/1; 32567/1; 32594/1; 32621/1; 32648/1; 32675/1; 32702/1; 32729/1; 32756/1; 32783/1; 32810/1; 32837/1; 32864/1; 32891/1; 32918/1; 32945/1; 32972/1; 33000/1; 33027/1; 33054/1; 33081/1; 33108/1; 33135/1; 33162/1; 33189/1; 33216/1; 33243/1; 33270/1; 33297/1; 33324/1; 33351/1; 33378/1; 33405/1; 33432/1; 33459/1; 33486/1; 33513/1; 33540/1; 33567/1; 33594/1; 33621/1; 33648/1; 33675/1; 33702/1; 33729/1; 33756/1; 33783/1; 33810/1; 33837/1; 33864/1; 33891/1; 33918/1; 33945/1; 33972/1; 34000/1; 34027/1; 34054/1; 34081/1; 34108/1; 34135/1; 34162/1; 34189/1; 34216/1; 34243/1; 34270/1; 34297/1; 34324/1; 34351/1; 34378/1; 34405/1; 34432/1; 34459/1; 34486/1; 34513/1; 34540/1; 34567/1; 34594/1; 34621/1; 34648/1; 34675/1; 34702/1; 34729/1; 34756/1; 34783/1; 34810/1; 34837/1; 34864/1; 34891/1; 34918/1; 34945/1; 34972/1; 35000/1; 35027/1; 35054/1; 35081/1; 35108/1; 35135/1; 35162/1; 35189/1; 35216/1; 35243/1; 35270/1; 35297/1; 35324/1; 35351/1; 35378/1; 35405/1; 35432/1; 35459/1; 35486/1; 35513/1; 35540/1; 35567/1; 35594/1; 35621/1; 35648/1; 35675/1; 35702/1; 35729/1; 35756/1; 35783/1; 35810/1; 35837/1; 35864/1; 35891/1; 35918/1; 35945/1; 35972/1; 36000/1; 36027/1; 36054/1; 36081/1; 36108/1; 36135/1; 36162/1; 36189/1; 36216/1; 36243/1; 36270/1; 36297/1; 36324/1; 36351/1; 36378/1; 36405/1; 36432/1; 36459/1; 36486/1; 36513/1; 36540/1; 36567/1; 36594/1; 36621/1; 36648/1; 36675/1; 36702/1; 36729/1; 36756/1; 36783/1; 36810/1; 36837/1; 36864/1; 36891/1; 36918/1; 36945/1; 36972/1; 37000/1; 37027/1; 37054/1; 37081/1; 37108/1; 37135/1; 37162/1; 37189/1; 37216/1; 37243/1; 37270/1; 37297/1; 37324/1; 37351/1; 37378/1; 37405/1; 37432/1; 37459/1; 37486/1; 37513/1; 37540/1; 37567/1; 37594/1; 37621/1; 37648/1; 37675/1; 37702/1; 37729/1; 37756/1; 37783/1; 37810/1; 37837/1; 37864/1; 37891/1; 37918/1; 37945/1; 37972/1; 38000/1; 38027/1; 38054/1; 38081/1; 38108/1; 38135/1; 38162/1; 38189/1; 38

db 1587 T M P F L F C D V N N V C N Y A S P N D K S Y W L

T29350

Best Local Similarity 80.0%; Pred. No. 8.1e-09;  
Matches 20; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

1585 TMPFLPCDFNNVCNYASRNDKSYWL 1609

Search completed: April 5, 2004, 07:05:35  
Job time : 6.14528 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.1477 Seconds  
(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221b-37

Perfect score: 147

Sequence: 1 TYPFLPCNVNDVCFASRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	147	100.0	1670	1 CA34 HUMAN	Q01955 homo sapien
2	146	99.3	471	1 CA34 BOVIN	Q28084 bos taurus
3	141	95.9	1669	1 CA14 HUMAN	P02462 homo sapien
4	141	95.9	1669	1 CA14 MOUSE	P02463 mus musculus
5	139	94.6	754	1 CA54 CANFA	Q28247 canis famil
6	139	94.6	1685	1 CA54 HUMAN	P29400 homo sapien
7	135	85.0	1763	1 CA24 ASCSU	P27393 ascaris suu
8	117	79.6	1758	1 CA14 CAEEL	P17139 caenorhabdi
9	116	78.9	1758	1 CA24 CAEEL	P17140 caenorhabdi
10	115	78.2	1691	1 CA64 HUMAN	Q14031 homo sapien
11	112	76.2	1707	1 CA24 MOUSE	P08122 mus musculus
12	112	76.2	1712	1 CA24 HUMAN	P08572 homo sapien
13	103	70.1	623	1 CA44 RABIT	P55787 oryctolagus
14	103	70.1	1690	1 CA44 HUMAN	P53420 homo sapien
15	102	69.4	453	1 CA44 BOVIN	Q29442 bos taurus
16	93	63.3	1775	1 CA14 DROME	P08120 drosophila
17	51	34.7	333	1 AMR1 HUMAN	Q9Y4X0 homo sapien
18	51	34.7	344	1 AMR1 MOUSE	Q9Y4X5 mus musculus
19	50.5	34.4	663	1 WM02 CHICK	Q09611 gallus gall
20	49	33.3	397	1 YMF7 YEAST	Q04359 saccharomyc
21	48	32.7	308	1 META_SALTI	Q821W1 salmonella
22	48	32.7	308	1 META_SALTY	P37413 salmonella
23	48	32.7	380	1 NTG2 YEAST	Q08214 saccharomyc
24	47	32.0	186	1 RYGL HUMAN	Q9BW83 homo sapien
25	46.5	31.6	379	1 MURV PORGI	Q7MAW5 porphyron
26	46	31.0	356	1 CMO_EYRFU	Q51741 pyrococcus
27	45.5	31.0	1743	1 TAGC_BICDI	Q23868 dictyosteli
28	45	30.6	1095	1 C2S5_SACKL	Q02342 saccharomyc
29	44.5	30.3	704	1 OE66_NPVAC	Q00704 autographa
30	44	29.9	260	1 NMA HUMAN	Q13145 homo sapien
31	44	29.9	384	1 Y092_RICPR	Q17917 caenorhabdi
32	43.5	29.6	334	1 Y092_RICPR	Q9ZE55 rickettsia
33	43.5	29.6	433	1 TCO1_HUMAN	P20061 homo sapien

## ALIGNMENTS

### RESULT 1

CA34 HUMAN STANDARD; PRT; 1670 AA.  
ID CA34 HUMAN STANDARD; PRT; 1670 AA.  
AC Q01955; Q9BOT2; 34, Created  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN R1  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94384994; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RL J. Biol. Chem. 269:23013-23017 (1994).  
[2]  
RN R2  
RP REVISIONS.  
RA Leinonen A.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
RN R3  
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;  
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND  
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;  
PRO-574; GLU-1269 AND PRO-1474.  
RX MEDLINE=21064696; PubMed=1134955;  
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106 (2001).  
[4]  
RN R4  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015836; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:19780-19784 (1992).  
[5]  
RN R5  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=182840;  
RA Morrison K.E., Mariyama M., Yang-Peng T.L., Readers S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554 (1991).  
[6]  
RN R6  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RX TISSUE=Kidney;  
RL MEDLINE=92147878; PubMed=1737849;

34 43.5 29.6 662 1 MM02 MOUSE  
35 43.5 29.6 662 1 MM02 RAT  
36 43 29.3 128 1 CD59 HUMAN  
37 43 29.3 343 1 Z183 HUMAN  
38 43 29.3 385 1 CHEB BORBU  
39 43 29.3 395 1 NH10 CAEEL  
40 43 29.3 464 1 SYR2 COXBU  
41 43 29.3 849 1 SRK6 BRAOL  
42 42.5 28.9 361 1 ALR CORGL  
43 42.5 28.9 407 1 NKIR MOUSE  
44 42.5 28.9 407 1 NKIR RAT  
45 42.5 28.9 407 1 NKIR RAT

P33434 mus musculus  
P33436 rattus norv  
P13987 h cd59 glyc  
O15541 homo sapien  
Q45047 bozrelia bu  
P41999 caenorhabdi  
O8316 coxiella bu  
Q09092 brassica ol  
Q8reus corynebacte  
P25103 homo sapien  
P30548 mus musculus  
P14600 rattus norv





```
Query Match      100.0%; Score 147; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. NO. 6.1e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFLFCNVNDVCFASRNDYSYWL 25
    |||||
Db 1499 TMEFLFCNVNDVCFASRNDYSYWL 1523

RESULT 2
ID CA34 BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reeders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
    encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gurwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
    alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
    of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438293;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
    Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
    membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
    glomerular basement membranes (GBM), forming a 'chicken-wire'
    meshwork together with laminins, proteoglycans and entactin/
    nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
    alpha 6(IV), each of which can form a triple helix structure
    with 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
    domain (NC1) at their C-terminus, frequent interruptions of the
    G-X-Y repeats in the long central triple-helical domain (which may
    cause flexibility in the triple helix), and a short N-terminal
    triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
    unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
    are involved in inter- and intramolecular disulfide bonding. 12 of
    these, located in the NC1 domain, are conserved in all known type
```

RP SEQUENCE OF 1-943 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=89029471; PubMed=3311751;  
RA Brazel D., Oberbauer I., Diezinger H., Babel W., Glanville R.W.,  
RA Deutzmann R., Kuehn K.;  
RT "Completion of the amino acid sequence of the alpha 1 chain of human  
RT basement membrane collagen (type IV) reveals 21 non-triplet  
RT interruptions located within the collagenous domain.";  
RL Eur. J. Biochem. 168:529-536(1987).  
RN [4]  
RN SEQUENCE OF 28-243.  
RP MEDLINE=8604708; PubMed=4043082;  
RA Glanville R.W., Qian R.O., Siebold B., Risteli J., Kuehn K.;  
RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
RT region (7S domain) of the alpha 1 (IV) chain of human basement  
RT membrane collagen.";  
RL Eur. J. Biochem. 152:213-219(1985).  
RN [5]  
RN SEQUENCE OF 534-1447.  
RX MEDLINE=8503629; PubMed=6434307;  
RA Babel W., Glanville R.W.;  
RT "Structure of human-basement-membrane (type IV) collagen. Complete  
RT amino-acid sequence of a 914-residue-long pepsin fragment from the  
RT alpha 1(IV) chain.";  
RL Eur. J. Biochem. 143:545-556(1984).  
RN [6]  
RN SEQUENCE OF 1256-1669 FROM N.A.  
RP MEDLINE=85207819; PubMed=2581959;  
RA Phlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
RA Cheung M.-C., Prockop D.J., Boyd C.D.;  
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV  
RT procollagen reveal an unusual homology of amino acid sequences in two  
RT halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687(1985).  
RN [7]  
RN SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RA Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
RT procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
RN [8]  
RN SEQUENCE OF 1-28 FROM N.A.  
RP MEDLINE=89034231; PubMed=3182844;  
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
RT collagen are divergently encoded on opposite DNA strands and have an  
RT overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220(1988).  
RN [9]  
RN SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RT of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624(1988).  
CC -1- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -1- PTM: Lysines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
unit (Type IV) collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
EMBL; M26576; AAA53098.1; -  
EMBL; J04217; AAA53098.1; JOINED.  
DR EMBL; M26550; AAA53098.1; JOINED.  
DR EMBL; M26540; AAA53098.1; JOINED.  
DR EMBL; M26542; AAA53098.1; JOINED.  
DR EMBL; M26543; AAA53098.1; JOINED.  
DR EMBL; M26544; AAA53098.1; JOINED.  
DR EMBL; M26545; AAA53098.1; JOINED.  
DR EMBL; M26546; AAA53098.1; JOINED.  
DR EMBL; M26547; AAA53098.1; JOINED.  
DR EMBL; M26537; AAA53098.1; JOINED.  
DR EMBL; M26538; AAA53098.1; JOINED.  
DR EMBL; M26548; AAA53098.1; JOINED.  
DR EMBL; M26549; AAA53098.1; JOINED.  
DR EMBL; M26551; AAA53098.1; JOINED.  
DR EMBL; M26552; AAA53098.1; JOINED.  
DR EMBL; M26553; AAA53098.1; JOINED.  
DR EMBL; M26554; AAA53098.1; JOINED.  
DR EMBL; M26555; AAA53098.1; JOINED.  
DR EMBL; M26556; AAA53098.1; JOINED.  
DR EMBL; M26557; AAA53098.1; JOINED.  
DR EMBL; M26558; AAA53098.1; JOINED.  
DR EMBL; M26559; AAA53098.1; JOINED.  
DR EMBL; M26560; AAA53098.1; JOINED.  
DR EMBL; M26561; AAA53098.1; JOINED.  
DR EMBL; M26562; AAA53098.1; JOINED.  
DR EMBL; M26563; AAA53098.1; JOINED.  
DR EMBL; M26564; AAA53098.1; JOINED.  
DR EMBL; M26565; AAA53098.1; JOINED.  
DR EMBL; M26566; AAA53098.1; JOINED.  
DR EMBL; M26567; AAA53098.1; JOINED.  
DR EMBL; M26568; AAA53098.1; JOINED.  
DR EMBL; M26569; AAA53098.1; JOINED.  
DR EMBL; M26570; AAA53098.1; JOINED.  
DR EMBL; M26571; AAA53098.1; JOINED.  
DR EMBL; M26572; AAA53098.1; JOINED.  
DR EMBL; M26573; AAA53098.1; JOINED.  
DR EMBL; M26574; AAA53098.1; JOINED.  
DR EMBL; M26575; AAA53098.1; JOINED.  
DR EMBL; Y00706; CAA23075.1; -  
DR EMBL; X05561; CAA23075.1; -  
DR EMBL; M10940; AAA52006.1; -  
DR EMBL; M11315; AAA52042.1; -  
PIR; S16876; CGHUB.  
DR Genew; HGNC:2202; COL4A1.  
DR MM; 120130; -  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagnc\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; Clg\_helix; 6.  
DR ProDom; PD003923; Procollagnc4; 1.  
DR SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Basement membrane;  
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1441 1669 NON-HELICAL REGION (NC1).  
 FT CARBOHYD 126 136 N-LINKED (GLCNAC. . .).  
 FT DISULFID 1460 1551 OR 1548.  
 FT DISULFID 1493 1548 OR 1551.  
 FT DISULFID 1505 1511  
 FT DISULFID 1570 1665  
 FT DISULFID 1604 1662  
 FT DISULFID 1616 1622  
 FT CONFLICT 237 238 SG -> KE (IN REF. 4).  
 FT CONFLICT 241 241 G -> K (IN REF. 4).  
 FT CONFLICT 319 319 Q -> A (IN REF. 3).  
 FT CONFLICT 719 719 N -> D (IN REF. 5).  
 FT CONFLICT 837 837 D -> Y (IN REF. 5).  
 FT CONFLICT 842 842 K -> P (IN REF. 5).  
 FT CONFLICT 896 896 V -> W (IN REF. 2).  
 FT CONFLICT 904 904 E -> Q (IN REF. 5).  
 FT CONFLICT 914 914 S -> K (IN REF. 5).  
 FT CONFLICT 998 998 S -> K (IN REF. 5).  
 FT CONFLICT 1010 1010 K -> P (IN REF. 5).  
 FT CONFLICT 1012 1012 S -> K (IN REF. 5).  
 FT CONFLICT 1358 1358 E -> Q (IN REF. 5).  
 SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFB938A84 CRC64;  
 Query Match 95.9%; Score 141; DB 1; Length 1669;  
 Best Local Similarity 92.0%; Pred. No. 4.3e-12;  
 Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWPFFCNVNDVNCNFSRNDYSYWL 25  
 |||||:|||||  
 Db 1499 TWPFFCNVNDVNCNFSRNDYSYWL 1523  
 RESULT 4  
 ID CA14 MOUSE STANDARD; PRT; 1669 AA.  
 AC P02463;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89197932; PubMed=2703490;  
 RA Muthukumar G., Blumberg B., Kurkinen M.;  
 RT "The complete primary structure for the alpha 1-chain of mouse  
 collagen IV. Differential evolution of collagen IV domains.";  
 RL J. Biol. Chem. 264:6310-6317(1989).  
 RN [2]  
 RP SEQUENCE OF 1-1154 FROM N.A.  
 RX MEDLINE=88112221; PubMed=3338568;  
 RA Wood L., Theriault N., Vogeli G.;  
 RT "cDNA clones completing the nucleotide and derived amino acid  
 sequence of the alpha 1 chain of basement membrane (type IV) collagen  
 from mouse.";  
 RL FEBS Lett. 227:5-8(1988).  
 RN [3]  
 RP SEQUENCE OF 1149-1424 FROM N.A.  
 RX MEDLINE=86301886; PubMed=3755692;  
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;  
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a  
 synthetic oligodeoxynucleotide.";  
 RL Gene 43:301-304(1986).  
 RN [4]  
 RP SEQUENCE OF 1276-1669 FROM N.A.  
 RX MEDLINE=85127033; PubMed=2578961;  
 RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
 Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;  
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of  
 the alpha 1(IV) chain of basement membrane collagen as derived from  
 complementary DNA.";  
 RL Eur. J. Biochem. 147:217-224(1985).  
 RN [5]  
 RP SEQUENCE OF 1441-1669 FROM N.A.  
 RX MEDLINE=87250460; PubMed=3597383;  
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
 Saus J., Pihlajaniemi T.;  
 RT "Extensive homology between the carboxyl-terminal peptides of mouse  
 alpha 1(IV) and alpha 2(IV) collagen.";  
 RL J. Biol. Chem. 262:8496-8499(1987).  
 RN [6]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=86196099; PubMed=3009468;  
 RA Sakurai Y., Sullivan M., Yamada Y.;  
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
 collagen genes.";  
 RL J. Biol. Chem. 261:6654-6657(1986).  
 RN [7]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89066738; PubMed=3198626;  
 RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
 RT "Head-to-head arrangement of murine type IV collagen genes.";  
 RL J. Biol. Chem. 263:19274-19277(1988).  
 RN [8]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89071759; PubMed=3200851;  
 RA Burdello P.D., Martin G.R., Yamada Y.;  
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
 bidirectional promoter and a shared enhancer.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).  
 RN [9]  
 RP SEQUENCE OF 1-129 FROM N.A.  
 RX MEDLINE=88243724; PubMed=3379041;  
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;  
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
 collagen chain and the corresponding region of the gene.";  
 RL J. Biol. Chem. 263:8706-8709(1988).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 glomerular basement membrane (GBM), forming a 'chicken-wire'  
 meshwork together with laminins, proteoglycans and entactin/  
 nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
 alpha 6(IV), each of which can form a triple helix structure with  
 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 domain (NC1) at their C-terminus, frequent interruptions of the G-  
 X-Y repeats in the long central triple-helical domain (which may  
 cause flexibility in the triple helix), and a short N-terminal  
 triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 are involved in inter- and intramolecular disulfide bonding. 12 of  
 these, located in the NC1 domain, are conserved in all known type  
 IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J03758; AAA37439.1; -.  
 CC EMBL; M23333; AAA51625.1; -.

```
DR EMBL; J04694; AAA50292.1; -
DR EMBL; X06777; CAA29946.1; -
DR EMBL; X02201; CAA26132.1; -
DR EMBL; M15832; AAA37340.1; -
DR EMBL; M14042; AAA37342.1; -
DR EMBL; M12879; AAA37343.1; -
DR EMBL; M13024; -; NOT_ANNOTATED_CDS.
DR EMBL; M13025; -; NOT_ANNOTATED_CDS.
DR EMBL; M13026; AAA37344.1; -
DR EMBL; M13027; AAA37345.1; -
DR EMBL; M13043; AAA37346.1; -
DR EMBL; J04448; AAA37437.1; -
DR PIR; A33525; CGMS4B.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:Basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM0011; C4; 2.
DR KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1662 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> P (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 3).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 95.9%; Score 141; DB 1; Length 1669;
Best Local Similarity 92.0%; Pred. No. 4.3e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 1499 TMPFLFCNVNDVCFASRNDYSYWL 1523

RESULT 5
ID_CA54_CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorner P.S., Marrano P., Bauml R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
human X-linked hereditary nephritis resulting from a single base
mutation in the gene encoding the alpha 5 chain of collagen type
IV."
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
canine X-linked hereditary nephritis (HN), a disease similar to
that in humans (also referred to as Alport syndrome) characterized
by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U07889; AAB60258.1; -.
DR PIR; A55267; A55267.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM0011; C4; 2.
DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT DOMAIN 1 1 TRIPLE-HELICAL REGION.
FT DOMAIN 531 754 NONHELICAL REGION (NC1).
FT DISULFID 532 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 7 OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON_TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 94.6%; Score 139; DB 1; Length 754;
Best Local Similarity 88.0%; Pred. No. 3.7e-12;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 591 TMPFLFCNVNDVCFASRNDYSYWL 615
```

## RESULT 6

CA54\_HUMAN STANDARD; PRT; 1685 AA.  
 ID CA54\_HUMAN STANDARD; PRT; 1685 AA.  
 AC P29400; Q16006; Q16126;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 5(IV) chain precursor.  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165049; PubMed=8120014;  
 RA Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RL J. Biol. Chem. 269:6608-6614(1994).  
 RN [2]  
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92316923; PubMed=1352287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 RN [3]  
 RP SEQUENCE OF 85-1685 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90337930; PubMed=2380186;  
 RA Pihlajaniemi T., Pohjola E.R., Myers J.C.;  
 RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";  
 RL J. Biol. Chem. 265:13758-13766(1990).  
 RN [4]  
 RP SEQUENCE OF 924-1685 FROM N.A.  
 RX MEDLINE=91169491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 RN [5]  
 RP SEQUENCE OF 914-1685 FROM N.A.  
 RX MEDLINE=90160375; PubMed=1689491;  
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;  
 RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6]  
 RP SEQUENCE OF 1442-1471 FROM N.A.  
 RX MEDLINE=90252791; PubMed=2339699;  
 RA Myers J.C., Jones T.A., Pohjola E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;  
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";  
 RL Am. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).  
 RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;  
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Sheets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [10]  
 RP VARIANT AS SER-1564.  
 RX MEDLINE=91169492; PubMed=1672282;  
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;  
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]  
 RP VARIANT AS ARG-325.  
 RX MEDLINE=92303559; PubMed=1376965;  
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";  
 RL Am. J. Hum. Genet. 51:135-142(1992).  
 RN [12]  
 RP VARIANT AS GLU-325.  
 RX MEDLINE=93244772; PubMed=1363780;  
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";  
 RL Hum. Mol. Genet. 1:127-129(1992).  
 RN [13]  
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
 RX MEDLINE=94010948; PubMed=8406498;  
 RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796; ARG-869; ARG-872 AND CYS-1241.  
 RX MEDLINE=95322976; PubMed=7599631;  
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the COL4A5 gene in Alport syndrome patients.";  
 RL Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=96213750; PubMed=8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M., Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is common in the western United States.";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS.  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S., Turco A.B., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G., Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F., Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51 exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).



```
FT isoform II)
FT /FTID=VSP_001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F52BEC06A80D CRC64;

Query Match 85.0%; Score 125; DB 1; Length 1763;
Best Local Similarity 84.0%; Pred. No. 8.4e-10;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 TMPPFLFCNVNDVCFNSASNDYSYL 25
DB 1587 TMPPFLFCNVNDVCFNSASNDYSYL 1611

RESULT 8
CA14_CABEL
ID CA14_CABEL STANDARD; PRT; 1758 AA.
AC FT1739;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CIB-2 OR KO4H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsys T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hallier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laester N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierly-Wieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlschlag P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.
CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
intermolecular interactions between 7S domains and between NC1
domains.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
```







```

DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U46999; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -
DR EMBL; AL109943; CAB88263.1; -
DR EMBL; AL136080; CAB96748.1; -
DR EMBL; AL031177; CAA20120.1; -
DR EMBL; L22763; AAA16338.1; -
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -
DR GO; GO:000587; C:collagen type IV; NAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01331; Collagen; 23.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6 (IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 78.2%; Score 115; DB 1; Length 1691;
Best Local Similarity 68.0%; Pred. No. 2.le-08;
Matches 17; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TMPLFCVNDVNCVFASRNDYSWL 25
Db 1521 TMPLFYCNINEVCHYARRNDKSYWL 1545

RESULT 11
CA24_MOUSE STANDARD; PRT; 1707 AA.
ID CA24_MOUSE
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2 (IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumaran G.,
RA Pihlajaniemi T., Kurkinen M.,

```

```

RT "The complete primary structure of mouse alpha 2 (IV) collagen.
RT Alignment with mouse alpha 1 (IV) collagen.";
RL J. Biol. Chem. 264:6318-6324 (1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277 (1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
RA Deutmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2 (IV)
RL chain and its comparison with the alpha 1 (IV) chain.";
RL Eur. J. Biochem. 157:49-56 (1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;
RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1 (IV).";
RL FEBS Lett. 208:203-207 (1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1 (IV) and alpha 2 (IV) collagen.";
RL J. Biol. Chem. 262:8496-8499 (1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:29-32 (1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3639908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RL alpha 2 (IV) collagen gene.";
RL Nature 317:177-179 (1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdeto P.D., Martin G.R., Yamada Y.;
RT "Alpha 1 (IV) and alpha 2 (IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1 (IV) -
CC alpha 6 (IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M23334; AAA51626.1; -;  
 DR EMBL; M23333; AAA51626.1; JOINED.  
 DR EMBL; J04695; AAA50293.1; -;  
 DR EMBL; J04448; AAA37438.1; -;  
 DR EMBL; X04647; CAA28308.1; -;  
 DR EMBL; M15833; AAA37341.1; -;  
 DR EMBL; X04410; CAA27998.1; -;  
 DR EMBL; X02896; CAA26655.1; -;  
 DR EMBL; X02897; CAA31644.1; -;  
 DR EMBL; X02898; CAA32657.1; -;  
 DR EMBL; X02899; CAA32658.1; -;  
 DR PIR; A33526; A33526.  
 DR MGI; MGI:89455; Col4a2.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.  
 DR ProDom; PD000007; C1g\_helix; 7.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT CHAIN 184 1707 TRIPLE-HELICAL REGION.  
 FT DOMAIN 184 1479 NONHELICAL REGION (NCL).  
 FT DOMAIN 1480 1707 OR 1588 (BY SIMILARITY).  
 FT DISULFID 1499 1588 OR 1588 (BY SIMILARITY).  
 FT DISULFID 1532 1585 BY SIMILARITY.  
 FT DISULFID 1544 1550 OR 1700 (BY SIMILARITY).  
 FT DISULFID 1607 1703 OR 1703 (BY SIMILARITY).  
 FT DISULFID 1641 1700 BY SIMILARITY.  
 FT DISULFID 1653 1660 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 1051 1051 P -> R (IN REF. 6).  
 FT CONFLICT 1097 1097 S -> G (IN REF. 7).  
 FT CONFLICT 1171 1171 G -> S (IN REF. 6).  
 FT CONFLICT 1179 1179 P -> R (IN REF. 6).  
 FT CONFLICT 1241 1241 Q -> E (IN REF. 6).  
 FT CONFLICT 1328 1328 P -> A (IN REF. 6).  
 FT CONFLICT 1573 1573 V -> L (IN REF. 4).  
 FT CONFLICT 1623 1623 Y -> H (IN REF. 4).  
 SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605PD508 CRC64;

Query Match 76.2%; Score 112; DB 1; Length 1707;  
 Best Local Similarity 76.0%; Pred. No. 5.7e-08;  
 Matches 19; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPEFLCNVDNCFASRDNSYWL 25  
 DB 1538 TMPEFLCNPGDVCYASRDNSYWL 1562

RESULT 12  
 CA24 HUMAN STANDARD; PRT; 1712 AA.  
 AC P08572;

DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 GN COL4A2.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89066769; PubMed=3198637;  
 RA Hostikka S.L., Tryggvason K.;  
 RT "The complete primary structure of the alpha 2 chain of human type IV  
 RT collagen and comparison with the alpha 1(IV) chain.";  
 RT J. Biol. Chem. 263:19488-19493(1988).  
 RN [2]  
 RP SEQUENCE OF 1-1042 FROM N.A.  
 RX TISSUE=Placenta;  
 MEDLINE=88151998; PubMed=3345760;  
 RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;  
 RT "Human basement membrane collagen (type IV). The amino acid sequence  
 RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)  
 RT chain reveals deletions in the alpha 1(IV) chain.";  
 RT Eur. J. Biochem. 172:35-42(1988).  
 RN [3]  
 RP SEQUENCE OF 1254-1712 FROM N.A.  
 RX MEDLINE=87219158; PubMed=3582677;  
 RA Hostikka S.L., Kurkinen M., Tryggvason K.;  
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2  
 RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1  
 RT (IV) but not with the collagenous domain or 3'-untranslated region.";  
 RT FEBS Lett. 216:281-286(1987).  
 RN [4]  
 RP SEQUENCE OF 1451-1485 FROM N.A.  
 RX MEDLINE=87092438; PubMed=3025878;  
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;  
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and  
 RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).  
 RN [5]  
 RP SEQUENCE OF 1486-1712 FROM N.A.  
 RX MEDLINE=87250571; PubMed=2439508;  
 RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;  
 RT "Duplication of type IV collagen COOH-terminal repeats and species-  
 RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";  
 RT J. Biol. Chem. 262:9231-9238(1987).  
 RN [6]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89034231; PubMed=3182844;  
 RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
 RT collagen are divergently encoded on opposite DNA strands and have an  
 RT overlapping promoter region.";  
 RT J. Biol. Chem. 263:17217-17220(1988).  
 RN [7]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89030632; PubMed=2846280;  
 RA Poeschl E., Pollner R., Kuehn K.;  
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human  
 RT basement membrane collagen type IV are arranged head-to-head and  
 RT separated by a bidirectional promoter of unique structure.";  
 RT EMBO J. 7:2667-2695(1988).  
 RN [8]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX TISSUE=Skin;  
 MEDLINE=93305049; PubMed=8317999;  
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;  
 RT "Identification of a novel sequence element in the common promoter  
 RT region of human collagen type IV genes, involved in the regulation of  
 RT divergent transcription.";  
 RT Biochem. J. 292:687-695(1993).  
 RN [9]  
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.  
 RC TISSUE=Placenta;  
 RX MEDLINE=89005112; PubMed=2844531;  
 RA Siebold B., Deutzmann R., Kuehn K.;  
 RT "The arrangement of intra- and intermolecular disulfide bonds in the

RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RL of basement-membrane type IV collagen.";  
CC Eur. J. Biochem. 176:617-624(1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM) forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05562; CAA29076.1; -  
CC EMBL; X05610; CAA29098.1; -  
CC EMBL; J02760; AAS58422.1; -  
CC EMBL; M36963; AAS30399.1; -  
CC EMBL; X12784; CAA31275.1; -  
CC EMBL; J04217; AAS30397.1; -  
CC PIR; A32024; CGH72B.  
CC Genew; HGNC:2203; COL4A2.  
CC MIM; 120090; -  
CC GO; GO:0005587; C:collagen type IV; TAS.  
CC GO; GO:0003201; F:extracellular matrix structural constituent; TAS.  
CC GO; GO:0030198; P:extracellular matrix organization and biogenesis; NAS.  
CC InterPro; IPR008161; Clg\_helix.  
CC InterPro; IPR008160; Collagen.  
CC InterPro; IPR001442; Procollagn\_4\_C.  
CC Pfam; PF01413; C4; 2.  
CC Pfam; PF01391; Collagen; 24.  
CC ProDom; PD000007; Clg\_helix; 7.  
CC ProDom; PD003923; ProcollagnC4; 1.  
CC SMART; SM00111; C4; 2.  
CC KW Glycoprotein; Basement membrane; Collagen; Signal.  
CC FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
CC FT PROPEP 26 183 COLLAGEN ALPHA 2(IV) CHAIN.  
CC FT CHAIN 184 1712 TRIPLE-HELICAL REGION.  
CC FT DOMAIN 184 1484 NONHELIICAL REGION (NC1).  
CC FT DISULFID 1485 1593 OR 1590 (BY SIMILARITY).  
CC FT DISULFID 1537 1550 OR 1593 (BY SIMILARITY).  
CC FT DISULFID 1549 1555 BY SIMILARITY.  
CC FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).  
CC FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).  
CC FT DISULFID 1658 1665 BY SIMILARITY.  
CC FT CARBOHYD 138 138 N-LINKED (GLCNAC...).  
CC FT CONFLICT 471 471 R -> P (IN REF. 2).  
CC FT CONFLICT 683 683 A -> G (IN REF. 2).  
CC FT CONFLICT 1575 1575 M -> I (IN REF. 5).  
CC FT CONFLICT 1663 1663 G -> H (IN REF. 9).  
CC FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
CC FT SEQUENCE 1712 AA; 167535 MW; 2582A17B47890037 CRC64;  
CC  
CC Query Match 76.2%; Score 112; DB 1; Length 1712;  
CC Best Local Similarity 76.0%; Pred. No. 5.7e-08;

Matches 19; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TMPLFCNVNDYCNFASNDYSYWL 25  
DB 1543 TMPLFCNVNDYCNFASNDYSYWL 1567  
RESULT 13  
CA44\_RABIT  
ID CA44\_RABIT STANDARD; PRT; 623 AA.  
AC P55787;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Collagen alpha 4(IV) chain (Fragment).  
GN COL4A4.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Corneal endothelium;  
RX MEDLINE=93054733; PubMed=1429714;  
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;  
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
RT alpha 4 chain of basement membrane collagen type IV and assignment of  
RT the gene to the distal long arm of human chromosome 2";  
RL J. Biol. Chem. 267:23753-23758(1992).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L01477; -; NOT\_ANNOTATED\_CDS.  
CC PIR; A45137; A45137.  
CC InterPro; IPR008160; Collagen.  
CC InterPro; IPR001442; Procollagn\_4\_C.  
CC Pfam; PF01413; C4; 2.  
CC Pfam; PF01391; Collagen; 5.  
CC ProDom; PD003923; ProcollagnC4; 1.  
CC SMART; SM00111; C4; 2.  
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
CC FT NON\_TER 1 1  
CC FT DOMAIN <1 392 TRIPLE-HELICAL REGION.  
CC FT DOMAIN 393 623 NONHELIICAL REGION (NC1).  
CC FT DISULFID 413 502 OR 499 (BY SIMILARITY).  
CC FT DISULFID 446 499 OR 502 (BY SIMILARITY).







FT NON\_TER 1  
 FT DOMAIN <1 222  
 FT DISULFID 223 453 TRIPLE-HELICAL REGION.  
 FT DISULFID 243 332 NONHELICAL REGION (NCL).  
 FT DISULFID 276 329 OR 329 (BY SIMILARITY).  
 FT DISULFID 288 294 OR 332 (BY SIMILARITY).  
 FT DISULFID 351 449 BY SIMILARITY.  
 FT DISULFID 385 446 OR 446 (BY SIMILARITY).  
 FT DISULFID 397 404 OR 449 (BY SIMILARITY).  
 FT CONFLICT 219 219 BY SIMILARITY.  
 SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

Query Match 69.4%; Score 102; DB 1; Length 453;  
 Best Local Similarity 60.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNVCNPNASNDYSYWL 25  
 DB 282 TLFFAYCNIHQVCHYARRNDRSYWL 306

Search completed: April 5, 2004, 06:59:38  
 Job time : 5.1477 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 15.0121 Seconds  
(without alignments) 525.440 Million cell updates/sec

Title: US-10-032-221B-37  
 Perfect score: 147  
 Sequence: 1 TMPFLFCNVNDVCFASRNDYSYWL 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
SPRENWBL_25.*
1: sp_archaea:
2: sp_bacteri:
3: sp_fungi:*
4: sp_human:*
5: sp_inverteb:
6: sp_mammal:
7: sp_mbc:*
8: sp_organel:
9: sp_phage:*
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_verteb:
14: sp_unclass:
15: sp_rvirus
16: sp_bacter:
17: sp_archaeu
18: sp_mammal
```

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	147	100.0	212	6	Q28512	macaca mula
2	147	100.0	245	4	Q9NYC4	homo sapien
3	146	99.3	203	6	Q29032	sus scrofa
4	146	99.3	203	6	Q28682	oryctolagus
5	146	99.3	212	6	Q28567	ovis aries
6	141	95.9	161	11	Q61430	mus musculus
7	141	95.9	210	6	Q28273	canis familiaris
8	141	95.9	225	6	Q28271	canis familiaris
9	141	95.9	226	11	Q99LQ8	mus musculus
10	141	95.9	229	4	Q9NF88	homo sapien
11	141	95.9	229	4	Q9NYC5	homo sapien
12	141	95.9	246	11	Q61435	mus musculus
13	141	95.9	979	13	Q919K3	gallus gallus
14	141	95.9	1075	4	Q86X41	homo sapien
15	141	95.9	1621	4	Q9H489	homo sapien
16	141	95.9	1669	11	Q9QZ50	mus musculus

## ALIGNMENTS

## RESULT 1

```

Q28512      PRELIMINARY;      PRT;      212 AA.
ID
Q28512;
AC
Q28512;
DT
01-NOV-1996 (TrEMBLrel. 01, Created)
DT
01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE
Alpha-3 type IV collagen (Fragment).
DD
COL4A3.
OS
Macaca mulatta (Rhesus macaque).
OC
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC
Cercopithecinae; Macaca.
OX
NCBI_TaxID=9544;
OX
[1]
RN
RN
RP
SEQUENCE FROM N.A.
RC
TISSUE=Kidney cortex;
RA
Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA
Mason P.J., Fusey C.D.;
RT
"Properties and sequences of the Goodpasture antigen of different
RT
mammals.";
RRL
Submitted (MAR-1996) to the EMBL/GenBank/DBSJ databases.
DR
EMBL; L47280; AAA91861.1; -.
DR
GO; GO:0005581; C:collagen; IEA.
DR
GO; GO:0005201; C:collagen; IEA.
DR
GO; GO:0003676; F:nucleic acid binding; IEA.
DR
InterPro; IPR01442; ProcollagN4 C.
DR
InterPro; IPR000504; RNA_rec_mot.
DR
Pfam; PF01413; C4; 2.
DR
ProDom; PD003923; ProcollagN4; 1.
DR
SMART; SMO0111; C4; 2.
DR
PROSITE; PS00030; RRM_RNP_1; 1.
KW
Collagen.
FT
NON_TER
FT
NON_TER
FT
NON_TER
FT
NON_TER
SQ
SEQUENCE      212 AA;      23469 MW;      4BC574A64E357B64      CRC64;
Q28512
Query Match      100.0%;      Score 147;      DB 6;      Length 212;
Best Local Similarity      100.0%;      Pred. No. 1.4e-14;

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25  
 |||||  
 DB 41 TMFFLCNVNDVCFASRNDYSYL 65

## RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.  
 AC Q9NYC4  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Tmsratin (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
 RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.;  
 RT "Distinct anti-tumor properties of a type IV collagen domain derived  
 RT from basement membrane.";  
 RL J. Biol. Chem. 0:0-0(2000).  
 DR EMBL; AF258351; AAP72632.1; -  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001442; Procollagn4 C.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 147; DB 4; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25  
 |||||  
 DB 74 TMFFLCNVNDVCFASRNDYSYL 98

## RESULT 3

Q29032 PRELIMINARY; PRT; 203 AA.  
 AC Q29032  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).  
 GN COL4A3.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
 RA Mason P.J., Pusey C.D.;  
 RT "Properties and sequences of the Goodpasture antigen of different  
 RT mammals.";  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L47284; AAA91882.1; -  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.

InterPro; IPR001442; Procollagn4 C.  
 InterPro; IPR000504; RNA\_rec\_mot.  
 Pfam; PF01413; C4; 2.  
 ProDom; PD003923; ProcollagnC4; 1.  
 SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Collagen.  
 FT NON\_TER 1  
 FT NON\_TER 203  
 SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 99.3%; Score 146; DB 6; Length 203;  
 Best Local Similarity 96.0%; Pred. No. 2e-14;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25  
 |||||  
 DB 41 TMFFLCNVNDVCFASRNDYSYL 65

## RESULT 4

Q28682 PRELIMINARY; PRT; 203 AA.  
 AC Q28682  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).  
 GN COL4A3.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
 RA Mason P.J., Pusey C.D.;  
 RT "Properties and sequences of the Goodpasture antigen of different  
 RT mammals.";  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L47283; AAA91893.1; -  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001442; Procollagn4 C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Collagen.  
 FT NON\_TER 1  
 FT NON\_TER 203  
 SQ SEQUENCE 203 AA; 22213 MW; E14173816B4D9E30 CRC64;

Query Match 99.3%; Score 146; DB 6; Length 203;  
 Best Local Similarity 96.0%; Pred. No. 2e-14;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFFLCNVNDVCFASRNDYSYL 25  
 |||||  
 DB 41 TMFFLCNVNDVCFASRNDYSYL 65

## RESULT 5

Q28567 PRELIMINARY; PRT; 212 AA.  
 AC Q28567  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).

```
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
CX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamna I.,
RA Mason P.J., Fusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; ProcollagN4 C.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 212 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 99.3%; Score 146; DB 6; Length 212;
Best Local Similarity 96.0%; Pred. No. 2.1e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 41 TMPFLFCNVNDVNCNFASRNDYSYWL 65

RESULT 6
Q61430 Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the NC1 domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (iv) and
RT alpha5 (iv) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X82205; CAA57689.1; -.
DR PIR: S49488; S49488.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 161 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE9236C5 CRC64;
```

```
Query Match 95.9%; Score 141; DB 11; Length 161;
Best Local Similarity 92.0%; Pred. No. 9.1e-14;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 8 TMPFLFCNVNDVNCNFASRNDYSYWL 32

RESULT 7
Q28273 Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A3 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL: U50935; AAC48585.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; ProcollagN4 C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR ProDom: PD003923; ProcollagN4 C.
DR SMART: SM00111; C4; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 1 1
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 95.9%; Score 141; DB 6; Length 210;
Best Local Similarity 92.0%; Pred. No. 1.2e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 51 TMPFLFCNVNDVNCNFASRNDYSYWL 75

RESULT 8
Q28271 Q28271 PRELIMINARY; PRT; 225 AA.
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
```

RT of collagen type IV. Evidence from a canine model of X-linked  
ST nephritis with a COL4A5 gene mutation.";  
RL J. Biol. Chem. 271:13821-13828(1996).  
RN [2]

RC SEQUENCE FROM N.A.

RA STRAIN=Samoyed;

RL Thorne P.S.;

RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: U50933; AAC48583.2; -

DR GO: GO:0005581; C:collagen; IEA.

DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.

DR InterPro: IPR001442; Procollagn4\_C.

DR Pfam: PF01413; C4; 2.

DR ProDom: PD003923; ProcollagnC4; 2.

DR SMART: SMC0111; C4; 2.

DR Collagen.

FT NON\_TER 1

FT NON\_TER 225

SQ SEQUENCE 225 AA; 24585 MW; 2C20455890416E47 CRC64;

Query Match 95.9%; Score 141; DB 6; Length 225;

Best Local Similarity 92.0%; Pred.No. 1.3e-13;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNVCNFCASRNDYSYWL 25

Db 65 TMPFLFCNINNVNVCNFCASRNDYSYWL 89

RESULT 9

Q99LQ8

ID Q99LQ8 PRELIMINARY; PRT; 226 AA.

AC Q99LQ8

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein (Fragment).

GN COL4A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC002269; AA002269.1; -

DR MGI: 88454; Col4a1.

DR InterPro: IPR001442; C:basement membrane; IDA.

DR Pfam: PF01413; C4; 2.

DR ProDom: PD003923; ProcollagnC4; 1.

DR SMART: SMC0111; C4; 2.

DR Hypothetical protein.

FT NON\_TER 1

FT NON\_TER 226

SQ SEQUENCE 226 AA; 25042 MW; 4F7F0D5371181C21 CRC64;

Query Match 95.9%; Score 141; DB 11; Length 226;

Best Local Similarity 92.0%; Pred.No. 1.3e-13;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNVCNFCASRNDYSYWL 25

Db 56 TMPFLFCNINNVNVCNFCASRNDYSYWL 80

RESULT 10

Q9NF88

ID Q9NF88 PRELIMINARY; PRT; 229 AA.

AC Q9NF88

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Arresten (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.

RA He A.B.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF536207; AA097359.1; -

DR GO: GO:0005581; C:collagen; IEA.

DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.

DR InterPro: IPR001442; Procollagn4\_C.

DR Pfam: PF01413; C4; 2.

DR ProDom: PD003923; ProcollagnC4; 1.

DR SMART: SMC0111; C4; 2.

FT NON\_TER 1

FT NON\_TER 229

SQ SEQUENCE 229 AA; 25391 MW; 09B21FD5AB517E9E CRC64;

Query Match 95.9%; Score 141; DB 4; Length 229;

Best Local Similarity 92.0%; Pred.No. 1.3e-13;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNVCNFCASRNDYSYWL 25

Db 59 TMPFLFCNINNVNVCNFCASRNDYSYWL 83

RESULT 11

Q9NYC5

ID Q9NYC5 PRELIMINARY; PRT; 229 AA.

AC Q9NYC5

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Arresten (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,

RA Takahashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K.,

RA Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D.,

RA Weichselbaum R.R., Sukhtme V.P., Kalluri R.;

RT "Anti-angiogenic cues from vascular basement membrane collagen.";

RL Cancer Res. 0:0-0(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Fu J., Bai X., Wang W., Ruan C.;

RA "Arresten, a collagen-derived inhibitor of angiogenesis.";

RL Chung Hua Hsueh Yeh Hsueh Tea Chih 22:0-0(2001).

RN [3]

RP SEQUENCE FROM N.A.

RA Peng X., Yin B., Yuan J., Qiang B.;

RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;

RT "Molecular cloning and sequencing of human arresten gene.";

RL Zhonghua Shi Yan Wei Ke Za Zhi 19:46-47(2002).

RN [5]

RP SEQUENCE FROM N.A.

RA Song Z.F., Zheng Q.C.;

RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF258349; AA072630.1; -

DR EMBL: AF363672; AA053382.1; -

DR EMBL: AF400431; AA092480.1; -

DR EMBL: AY285780; AA043112.1; -

DR GO: GO:0005581; C:collagen; IEA.

```
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMC0111; C4; 2.
FT NON_TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5CID5 CRC64;

Query Match 95.9%; Score 141; DB 4; Length 229;
Best Local Similarity 92.0%; Pred. No. 1.3e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 59 TMPFLFCNVNDVNCNFASRNDYSYWL 83

RESULT 12
Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
laminae: Sequence, distribution, association with laminins, and
developmental switches.";
RL J. Cell Biol. 127:879-891 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H.;
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z35166; CAA84529.1; -.
DR PIR; I48302; I48302.
DR MGD; MGI:104689; Ccl4a3.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMC0111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 95.9%; Score 141; DB 11; Length 246;
Best Local Similarity 92.0%; Pred. No. 1.4e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 75 TMPFLFCNVNDVNCNFASRNDYSYWL 99

RESULT 13
Q919K3 PRELIMINARY; PRT; 979 AA.
AC Q919K3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Collagen IV alpha 3 chain (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
basal lamina.";
RL Dev. Biol. 0:0-0 (2000).
DR EMBL; AF239838; AAF44681.1; -.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 12.
DR ProDom; PD000007; C1g_helix; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 95.9%; Score 141; DB 13; Length 979;
Best Local Similarity 92.0%; Pred. No. 5.8e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
Db 809 TMPFLFCNVNDVNCNFASRNDYSYWL 833

RESULT 14
Q86X41 PRELIMINARY; PRT; 1075 AA.
AC Q86X41;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 95.9%; Score 141; DB 4; Length 1075;
Best Local Similarity 92.0%; Pred. No. 6.4e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASRNDYSYWL 25
```

Search completed: April 5, 2004, 07:03:57  
Job time : 15.0121 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 / Search time 22.5182 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221b-37

Perfect score: 147

Sequence: 1 TWPFLLFCNVNDVCFASRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	100.0	25	ADA20236	Ada20236 T7 peptid
2	147	100.0	79	Aau75600	Aau75600 Human typ
3	147	100.0	79	ADA20264	Ada20264 Human tum
4	147	100.0	88	Aau75608	Aau75608 Human typ
5	147	100.0	88	Aau75607	Aau75607 Human typ
6	147	100.0	88	ADA20271	Ada20271 Human tum
7	147	100.0	88	ADA20272	Ada20272 Human tum
8	147	100.0	124	Aau75594	Aau75594 Human typ
9	147	100.0	124	ADA20258	Ada20258 Human tum
10	147	100.0	132	Aau75597	Aau75597 Human typ
11	147	100.0	132	ADA20261	Ada20261 Human tum
12	147	100.0	191	Aau75596	Aau75596 Human typ
13	147	100.0	191	ADA20260	Ada20260 Human tum
14	147	100.0	211	AAY95918	Aay95918 Human Goo
15	147	100.0	211	ABG79208	Abg79208 Human GP
16	147	100.0	218	AAR79164	Aar79164 Partial s
17	147	100.0	218	AAY44172	Aay44172 Human typ
18	147	100.0	218	AAY56784	Aay56784 Human alp
19	147	100.0	218	AAE09484	Aae09484 Human alp
20	147	100.0	232	ADC17697	Adc17697 Human typ
21	147	100.0	244	ABG79218	Abg79218 Human Goo
22	147	100.0	244	ABG79219	Abg79219 Human Goo
23	147	100.0	244	ABG79217	Abg79217 Human typ
24	147	100.0	244	Aau75595	Aau75595 Human typ
25	147	100.0	244	ADA20225	Ada20225 Human typ

26	147	100.0	245	3	RAY67942	Ray67942 Human typ
27	147	100.0	245	5	AAU75589	Aau75589 Human typ
28	147	100.0	254	5	AAU75598	Aau75598 Human typ
29	147	100.0	268	2	AAI31993	Aay31993 Type IV c
30	147	100.0	268	3	AAI31993	Aay31993 Type IV c
31	147	100.0	1670	7	ADD47063	Add47063 Human pro
32	146	99.3	471	2	AAI79163	Aar79163 Partial s
33	146	99.3	471	2	AAI79163	Aar79163 Partial s
34	146	99.3	471	3	AAI44171	Aay44171 Bovine ty
35	146	99.3	471	3	AAI56783	Aay56783 Bovine al
36	146	99.3	471	4	AAE09483	Aae09483 Bovine al
37	141	95.9	229	1	AAI31524	Aap31524 Complete
38	141	95.9	229	3	AAI67943	Aay67943 Human typ
39	141	95.9	229	5	AAU75587	Aau75587 Human typ
40	141	95.9	229	6	ADA20217	Ada20217 Human typ
41	141	95.9	229	7	ADC17695	Adc17695 Human typ
42	141	95.9	260	2	AAI31991	Aay31991 Type IV c
43	141	95.9	260	3	AAI31991	Aay31991 Type IV c
44	141	95.9	406	3	AAI58169	Aab58169 Lung.canc
45	141	95.9	1669	4	AAI40863	Aam40863 Human pol
	141	95.9	1669	5	ABB90760	Abb90760 Human Tum

## ALIGNMENTS

### RESULT 1

ADA20236  
ID ADA20236 standard; peptide; 25 AA.

XX AC ADA20236;

XX DT 20-NOV-2003 (first entry)

XX DE T7 peptide related to human type IV collagen alpha and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX KW metastasis; basement membrane organisation; type IV collagen network;

XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX KW cytosolic; gene therapy; T7 peptide; tumstatin; human;

XX KW type IV collagen alpha 3 chain.

XX OS Homo sapiens.

XX FN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX PS WPI; 2003-587256/55.

XX DR N-ESDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 53; Page 45; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments

XX CC with anti-angiogenic properties. The invention also relates to the DNA

XX CC sequences which encode the novel proteins. A wide variety of diseases are

XX CC the result of undesirable angiogenesis. The formation of new capillaries

XX CC from pre-existing vessels is essential for tumour growth and metastasis.

XX CC Basement membrane organisation is dependent on the assembly of a type IV

XX CC collagenous network which may occur through the C-terminal globular non-

XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

XX CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the T7 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.  
 CC  
 SQ Sequence 25 AA;

Query Match 100.0%; Score 147; DB 6; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCNFAFRNDYSYWL 25  
 |||||  
 DB 1 TMPFLFCNVNDVCNFAFRNDYSYWL 25

RESULT 2  
 AAU75600  
 ID AAU75600 standard; protein; 79 AA.  
 XX  
 AC AAU75600;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tum-5.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 XX Tumstatin; angiogenesis; tumour; mutant.

OS Homo sapiens.

PN WO200151523-A2.

PD 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 XX treating disorders involving angiogenesis.

XX Example 40; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 XX domain, having one or more of the characteristics selected from: (a) the  
 XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 XX proliferation of endothelial cells; and (c) the ability to cause  
 XX apoptosis of endothelial cells. Also described are the following: (1) use  
 XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 XX analogue or allelic variant in the preparation of a medicament for  
 XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 XX where the angiogenesis is mediated by one or more endothelial cell  
 XX integrins or one or more endothelial cell integrin subunits; or (b) by  
 XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
 XX endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues  
 CC 54-132 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75589)

SQ Sequence 79 AA;

Query Match 100.0%; Score 147; DB 5; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVCNFAFRNDYSYWL 25  
 |||||  
 DB 21 TMPFLFCNVNDVCNFAFRNDYSYWL 45

RESULT 3

ADA20264

ID ADA20264 standard; protein; 79 AA.

XX

AC ADA20264;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-5 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX Homo sapiens.

OS WO2003059257-A2.

PN 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX  
 PS Claim 94; SEQ ID NO 26; 240pp; English.  
 XX

CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.  
 XX

SQ Sequence 79 AA;

Query Match 100.0%; Score 147; DB 6; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25  
 |||||  
 DB 20 TMAPFLFCNVNDVCFASRNDYSYWL 44  
 |||||

RESULT 4

AAU75608  
 ID AAU75608 standard; protein; 88 AA.

AC AAU75608;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 82 /note= "Wild type Cys substituted with Ala"

XX WC200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX

DR WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.

XX Claim 41; Page 153; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors or Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which  
 CC consists of residues 5-126 of Tumstatin  
 XX

SQ Sequence 88 AA;

Query Match 100.0%; Score 147; DB 5; Length 88;

Best Local Similarity 100.0%; Pred. No. 5.2e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMAPFLFCNVNDVCFASRNDYSYWL 25  
 |||||

DB 30 TMAPFLFCNVNDVCFASRNDYSYWL 54  
 |||||

RESULT 5

AAU75607

ID AAU75607 standard; protein; 88 AA.

XX AAU75607;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

OS

WO200151523-A2.

19-JUL-2001.

08-JAN-2001; 2001WO-US000565.

07-JAN-2000; 2000US-00479118.

04-APR-2000; 2000US-00543371.

21-JUL-2000; 2000US-00625191.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

Kalluri R;

WPI; 2002-188037/24.

A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and treating disorders involving angiogenesis.

Claim 32; Page 152; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists of residues 45-132 of Tumstatin

Sequence 88 AA;

ADA20271  
ID ADA20271 standard; protein; 88 AA.  
XX  
XX AC AC  
XX ADA20271;  
XX  
XX DT 20-NOV-2003 (first entry)  
XX  
XX DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.  
XX  
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NC1; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 45-132.  
XX  
XX OS Homo sapiens.  
XX  
XX WO2003059257-A2.  
XX  
XX PN 24-JUL-2003.  
XX  
XX PD 20-DEC-2002; 2002WO-US040938.  
XX  
XX PF 21-DEC-2001; 2001US-00032221.  
XX  
XX PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
XX PA Kalluri R;  
XX  
XX PI WPI; 2003-587256/55.  
XX  
XX DR N-PSDB; ADA20224.  
XX  
XX PT New peptide, useful for preparing a composition for inhibiting tumour growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX PS Claim 94; SEQ ID NO 33; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis. Basement membrane organisation is dependent on the assembly of a type IV collagen network which may occur through the C-terminal globular non-collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2 forms are ubiquitously exhibited in human basement membranes. In the present invention, cell surface receptors (in particular integrins) which specifically bind anti-angiogenic proteins and peptides (in particular the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV collagen) are disclosed. The proteins of the invention may inhibit tumour growth, angiogenic activity in mammalian tissue or protein synthesis in endothelial cells and thus may exhibit cytostatic activity. The DNA sequences of the invention may be useful in gene therapy. The present sequence is that of tumstatin 45-132, an abridged form of the "tumstatin" protein of the invention which was derived from the amino acid sequence of the alpha 3 chain of human type IV collagen. Note: this sequence (Seq ID33) does not appear in the specification but was created by the indexer from information given in the specification.  
XX  
XX Sequence 88 AA;  
XX

ADA20272 standard; protein; 88 AA.  
 ADA20272;  
 20-NOV-2003 (first entry)  
 Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.  
 anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NCI; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 5-125-C-A; mutant; mutein.  
 Synthetic.  
 Homo sapiens.  
 Key Location/Qualifiers  
 Misc-difference 81  
 /note= "Wild-type Cys substituted by Ala at position 125 of full-length tumstatin"  
 WO2003059257-A2.  
 24-JUL-2003.  
 20-DEC-2002; 2002WO-US040938.  
 21-DEC-2001; 2001US-00032221.  
 (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 Kalluri R;  
 WPI; 2003-587256/55.  
 New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 Claim 94; SEQ ID NO 34; 240pp; English.  
 This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis. Basement membrane organisation is dependent on the assembly of a type IV collagen network which may occur through the C-terminal globular non-collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2 forms are ubiquitously exhibited in human basement membranes. In the present invention, cell surface receptors (in particular integrins) which specifically bind anti-angiogenic proteins and peptides (in particular the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV collagen) are disclosed. The proteins of the invention may inhibit tumour growth, angiogenic activity in mammalian tissue or protein synthesis in endothelial cells and thus may exhibit cytostatic activity. The DNA sequences of the invention may be useful in gene therapy. The present invention is that of tumstatin 5-125-C-A, a mutated and abridged form of the "tumstatin" protein of the invention which was derived from the amino acid sequence of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq ID33) does not appear in the specification but was created by the indexer from information given in the specification.  
 Sequence 88 AA;

Query Match 100.0%; Score 147; DB 6; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 TWPFLFCNVNDVCFASRNDYSYL 25  
 29 TWPFLFCNVNDVCFASRNDYSYL 53

RESULT 8  
 AAU75594  
 ID AAU75594 standard; protein; 124 AA.  
 XX  
 AC AAU75594;  
 XX  
 DT 03-MAY-2002 (first entry)  
 XX  
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.  
 XX  
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphasbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; arresten; canstatin;  
 KW tumstatin; angiogenesis; tumour; mutein; mutant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151523-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 08-JAN-2001; 2001WO-US000565.  
 XX  
 PR 07-JAN-2000; 2000US-00479118.  
 PR 04-APR-2000; 2000US-00543371.  
 PR 21-JUL-2000; 2000US-00625191.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Kalluri R;  
 XX  
 DR WPI; 2002-188037/24.  
 XX  
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and treating disorders involving angiogenesis.  
 XX  
 PS Example 33; Page; 205pp; English.  
 XX  
 CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell apoptosis in a tissue, where the promoting or inducing endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell apoptosis; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments inducing angiogenesis or cell proliferation are Arresten, Canstatin or Tumstatin or their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell apoptosis; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of  
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in  
CC the specification but is derived from the wild type human Tumstatin  
CC sequence given in figure 18A (see AAU75589)

XX  
SQ Sequence 124 AA;

Query Match 100.0%; Score 147; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 7.5e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFLFCNVNDVCFASRNDYSYWL 25  
| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 73 TMEFLFCNVNDVCFASRNDYSYWL 97

RESULT 9

ADA20258  
ID ADA20258 standard; protein; 124 AA.

XX  
AC ADA20258;

XX  
DT 20-NOV-2003 (first entry)

XX  
DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX  
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytotatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

XX  
OS Homo sapiens.

XX  
PN WO2003059257-A2.

XX  
PD 24-JUL-2003.

XX  
PF 20-DEC-2002; 2002WO-US040938.

XX  
PR 21-DEC-2001; 2001US-00032221.

XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX  
PI Kalluri R;

XX  
WPI; 2003-587256/55.  
N-PSDB; ADA20224.

XX  
PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX  
PS Claim 94; SEQ ID NO 20; 240pp; English.

XX  
CC This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytotatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present

CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
CC ID20) does not appear in the specification but was created by the Indexer  
CC from information given in the specification.

XX  
SQ Sequence 124 AA;

Query Match 100.0%; Score 147; DB 6; Length 124;  
Best Local Similarity 100.0%; Pred. No. 7.5e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFLFCNVNDVCFASRNDYSYWL 25  
| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 73 TMEFLFCNVNDVCFASRNDYSYWL 97

RESULT 10

AAU75597  
ID AAU75597 standard; protein; 132 AA.

XX  
AC AAU75597;

XX  
DT 08-MAY-2002 (first entry)

XX  
DE Human type IV collagen alpha 3 chain mutant, Tum-2.

XX  
KW Human; type IV collagen alpha 3 chain; cytotatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphabeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX  
OS Homo sapiens.

XX  
PN WO200151523-A2.

XX  
PD 19-JUL-2001.

XX  
PF 08-JAN-2001; 2001WO-US000565.

XX  
PR 07-JAN-2000; 2000US-00479118.

XX  
PR 04-APR-2000; 2000US-00543371.

XX  
PR 21-JUL-2000; 2000US-00625191.

XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX  
PI Kalluri R;

XX  
WPI; 2002-188037/24.

XX  
PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.

XX  
PS Claim 31; Page 152; 205pp; English.

XX  
CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphabeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation

CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues  
CC 1-132 of Tumstatin. Note: The present sequence is not shown in the  
CC specification but is derived from the wild type human Tumstatin sequence  
CC given in figure 18A (see AAU75583)

XX Sequence 132 AA;

Query Match 100.0%; Score 147; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 8e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25  
Db 74 TMPFLFCNVNDVCFASRNDYSYWL 98  
|||||

RESULT 11

ID ADA20261 standard; protein; 132 AA.

XX ADA20261;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX cytosstatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor

XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of  
CC the invention which was derived from the amino acid sequence of the alpha  
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does  
CC not appear in the specification but was created by the indexer from  
CC information given in the specification.

XX Sequence 132 AA;

Query Match 100.0%; Score 147; DB 6; Length 132;  
Best Local Similarity 100.0%; Pred. No. 8e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25  
Db 73 TMPFLFCNVNDVCFASRNDYSYWL 97  
|||||

RESULT 12

AAU75596

ID AAU75596 standard; protein; 131 AA.

XX AAU75596;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.

XX Example 32; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit



CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of  
 CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in  
 CC the specification but is derived from the wild type human Tumstatin  
 CC sequence given in figure 18A (see AAU75589)

XX Sequence 191 AA;

Query Match 100.0%; Score 147; DB 5; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFLCNVNDVCFNFSRNDYSYWL 25  
 |||||  
 Db 21 TMAPFLCNVNDVCFNFSRNDYSYWL 45

RESULT 13  
 ADA20260  
 ID ADA20260 standard; protein; 191 AA.

XX AC ADA20260;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-1 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;  
 KW tumstatin N53.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX XX 24-JUL-2003.

XX PD 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.  
 XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI, 2003-587256/55.  
 DR N-FSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX Claim 94; SEQ ID NO 22; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-1 (tumstatin N53), an abridged form of the  
 CC "tumstatin" protein of the invention which was derived from the amino  
 CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
 CC sequence (Seq ID22) does not appear in the specification but was created  
 CC by the indexer from information given in the specification.

XX Sequence 191 AA;

Query Match 100.0%; Score 147; DB 6; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFLCNVNDVCFNFSRNDYSYWL 25  
 |||||  
 Db 20 TMAPFLCNVNDVCFNFSRNDYSYWL 44

RESULT 14  
 AAY95918  
 ID AAY95918 standard; protein; 211 AA.

XX AC AAY95918;

XX DT 20-NOV-2000 (first entry)

XX DE Human Goodpasture antigen Deltav.

XX Goodpasture antigen; GPdeltav; goodpasture antigen binding protein; GPBP;  
 KW human; autoimmune disease; apoptosis; cancer; tumour; therapy.

XX OS Homo sapiens.

XX PN WO2000050607-A2.

XX PD 31-AUG-2000.

XX PF 24-FEB-2000; 2000WO-IB000324.

XX PR 24-FEB-1999; 99US-0121483P.

XX PA (SAUS/) SAUS J.

```
PI Saus J;
XX WPI; 2000-572094/53.
XX N-PSDB; AAA50367.
XX Novel Goodpasture antigen binding proteins useful for diagnosing and
XX treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX Claim 36; Page 151-152; 158pp; English.
XX The present sequence is that of human recombinant Goodpasture antigen
XX (GP) Deltav, i.e. an alternative form of human GP resulting from splicing
XX out of exon V. The recombinant protein, lacking the Met-1 residue, was
XX expressed in bacterial pellets using modified vector pET15b carrying
XX GPDeltav cDNA (see AAA50367). The invention relates to novel Goodpasture
XX antigen binding proteins (GPBPs, see AAY55900-11), which bind to and
XX phosphorylate the unique N-terminal region of human GP, and which are
XX highly expressed in several autoimmune conditions. Claimed methods for
XX treating an autoimmune disorder, cell apoptosis or a tumour involve
XX modifying the expression or activity of GPBPs, especially using a GP-
XX derived peptide, such as GPDeltav
XX SQ Sequence 211 AA;
XX Query Match 100.0%; Score 147; DB 3; Length 211;
XX Best Local Similarity 100.0%; Pred. No. 1.3e-13;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFLFCNVNVCNPFASRNDYSYWL 25
DB 73 TMPFLFCNVNVCNPFASRNDYSYWL 97
RESULT 15
ABG79208
ID ABG79208 standard; protein; 211 AA.
XX AC ABG79208;
XX DT 15-NOV-2002 (first entry)
XX DE Human GP protein isoform GPDeltav.
XX KW Goodpasture antigen binding protein; Goodpasture syndrome;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX OS Homo sapiens.
XX PN WO200261430-A2.
XX PD 08-AUG-2002.
XX PF 31-JAN-2002; 2002WO-EP001010.
XX PR 31-JAN-2001; 2001US-0265249P.
XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX WPI; 2002-619280/66.
XX DR N-PSDB; ABS64491.
XX Identifying candidate compounds for treating autoimmune conditions, e.g.
XX Goodpasture syndrome or lupus, comprises identifying compounds that
XX reduce phosphorylation of, or formation of conformational isomers of,
XX target proteins.
XX Example 3; Page 199-200; 217pp; English.
```

```
XX The invention relates to identifying candidate compounds to treat an
XX autoimmune condition by identifying compounds that reduce phosphorylation
XX of a first target protein (I) (which is selected from Goodpasture antigen
XX binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)
XX domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-
XX Ala-Thr-Trip-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX conformational isomers of the second target protein (II) (selected from
XX an alpha3 type IV collagen NC1 domain polypeptide and myelin basic
XX protein, MBP). Also included are (1) an isolated type IV collagen alpha3
XX NC1 domain conformational isomer, which has an amino acid sequence
XX identical to the wild type alpha3 type IV collagen NC1 domain, is
XX stabilised by disulphide bonds, and has a molecular weight in a non-
XX reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX a reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX type IV collagen alpha3 NC1 domain. The human gene for GPBP is located on
XX chromosome 5q13. The method is useful for treating autoimmune conditions,
XX such as Goodpasture syndrome, multiple sclerosis, systemic and cutaneous
XX lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX sequence represents an alpha3 type IV collagen non-collagenous (NC1)
XX domain (also known as the GP antigen) or an MBP isoform
XX SQ Sequence 211 AA;
XX Query Match 100.0%; Score 147; DB 5; Length 211;
XX Best Local Similarity 100.0%; Pred. No. 1.3e-13;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TMPFLFCNVNVCNPFASRNDYSYWL 25
DB 73 TMPFLFCNVNVCNPFASRNDYSYWL 97
Search completed: April 5, 2004, 06:58:31
Job time : 22.5182 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 15.678 Seconds

(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-37

Perfect score: 147

Sequence: 1 TMPFFLCNVNDVCFNDRNDYSYL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	100.0	25	14	US-10-032-221B-37
2	147	100.0	79	14	US-10-032-221B-26
3	147	100.0	88	14	US-10-032-221B-33
4	147	100.0	88	14	US-10-032-221B-34
5	147	100.0	124	14	US-10-032-221B-20
6	147	100.0	132	14	US-10-032-221B-23
7	147	100.0	191	14	US-10-032-221B-22
8	147	100.0	211	14	US-10-270-877-46
9	147	100.0	211	14	US-10-270-837-46
10	147	100.0	232	14	US-10-206-699-304
11	147	100.0	244	14	US-10-032-221B-10
12	141	95.9	229	14	US-10-206-699-302
13	141	95.9	229	14	US-10-032-221B-2
14	141	95.9	406	9	US-09-925-302-507
15	141	95.9	1669	15	US-10-372-683-8

16	139	94.6	25	14	US-10-032-221B-38	Sequence 38, Appl
17	139	94.6	229	14	US-10-206-699-306	Sequence 306, App
18	139	94.6	309	9	US-03-925-237-436	Sequence 436, App
19	129	87.8	46	9	US-09-864-761-48095	Sequence 48095, A
20	125	85.0	27	14	US-10-032-221B-39	Sequence 39, Appl
21	120	81.6	1759	15	US-10-369-493-7032	Sequence 7032, Ap
22	117	79.6	1744	15	US-10-369-493-5832	Sequence 5832, Ap
23	115	78.2	142	9	US-09-864-761-38021	Sequence 38021, A
24	115	78.2	228	14	US-10-206-699-307	Sequence 307, App
25	115	77.6	227	14	US-10-206-699-266	Sequence 266, App
26	112	76.2	227	14	US-10-206-699-303	Sequence 303, App
27	112	76.2	227	14	US-10-032-221B-6	Sequence 6, Appli
28	112	76.2	430	9	US-09-925-302-518	Sequence 518, App
29	112	76.2	459	15	US-10-331-496A-27	Sequence 27, Appl
30	112	76.2	459	15	US-10-372-683-30	Sequence 30, Appl
31	112	76.2	1712	10	US-09-961-403-9	Sequence 9, Appli
32	108	73.5	22	14	US-10-206-699-265	Sequence 265, App
33	106	72.1	22	14	US-10-206-699-267	Sequence 267, App
34	105	71.4	27	14	US-10-032-221B-40	Sequence 40, Appl
35	104	70.7	18	14	US-10-206-699-260	Sequence 260, App
36	103	70.1	231	14	US-10-206-699-305	Sequence 305, App
37	101	68.7	27	14	US-10-032-221B-42	Sequence 42, Appl
38	98	66.7	18	14	US-10-206-699-259	Sequence 259, App
39	96	65.3	18	14	US-10-206-699-261	Sequence 261, App
40	93	63.3	18	14	US-10-206-699-254	Sequence 254, App
41	93	63.3	19	14	US-10-032-221B-41	Sequence 41, Appl
42	91	61.9	22	14	US-10-206-699-270	Sequence 270, App
43	89	60.5	15	14	US-10-206-699-212	Sequence 212, App
44	89	60.5	20	14	US-10-206-699-289	Sequence 289, App
45	89	60.5	20	14	US-10-032-221B-29	Sequence 29, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-032-221B-37  
; Sequence 37, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS OF USE THERE  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,669  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 37  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T7 (amino acids 73-97 of SEQ ID NO:10)  
US-10-032-221B-37

Query Match 100.0%; Score 147; DB 14; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.1e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25

## RESULT 2

US-10-032-221B-26  
; Sequence 26, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 26  
; LENGTH: 79  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)  
US-10-032-221B-26

Query Match 100.0%; Score 147; DB 14; Length 79;  
Best Local Similarity 100.0%; Pred. No. 6.7e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 20 TMPFLFCNVNDVNCNFASTRNDYSYWL 44

## RESULT 3

US-10-032-221B-33  
; Sequence 33, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58

Query Match 100.0%; Score 147; DB 14; Length 79;  
Best Local Similarity 100.0%; Pred. No. 6.7e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 20 TMPFLFCNVNDVNCNFASTRNDYSYWL 44

; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 33  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)  
US-10-032-221B-33

Query Match 100.0%; Score 147; DB 14; Length 88;  
Best Local Similarity 100.0%; Pred. No. 7.4e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 29 TMPFLFCNVNDVNCNFASTRNDYSYWL 53

## RESULT 4

US-10-032-221B-34  
; Sequence 34, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 34  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tumstatin-5-135-C-A (amino acids 45-132 of SEQ ID NO:10; alanine  
; OTHER INFORMATION: has been substituted for the cysteine residue at position 125 of  
; OTHER INFORMATION: the full-length Tumstatin molecule)  
US-10-032-221B-34

Query Match 100.0%; Score 147; DB 14; Length 88;  
Best Local Similarity 100.0%; Pred. No. 7.4e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 29 TMPFLFCNVNDVNCNFASTRNDYSYWL 53

## RESULT 5

US-10-032-221B-20  
; Sequence 20, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21

Query Match 100.0%; Score 147; DB 14; Length 88;  
Best Local Similarity 100.0%; Pred. No. 7.4e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFLFCNVNDVNCNFASTRNDYSYWL 25  
DB 29 TMPFLFCNVNDVNCNFASTRNDYSYWL 53

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 7

US-10-032-221B-22

Sequence 22, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 20

LENGTH: 124

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)

US-10-032-221B-20

Query Match 100.0%; Score 147; DB 14; Length 124;

Best Local Similarity 100.0%; Pred. No. 1e-13; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 6

US-10-032-221B-23

Sequence 23, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 23

LENGTH: 132

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)

US-10-032-221B-23

Query Match 100.0%; Score 147; DB 14; Length 132;

Best Local Similarity 100.0%; Pred. No. 1.1e-13; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 73 TMAPLFCVNDVCFASRNDYSYWL 97

RESULT 7

US-10-032-221B-22

Sequence 22, Application US/10032221B

Publication No. US20030144481A1

GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

CURRENT APPLICATION NUMBER: US/10/032,221B

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: PCT/US01/00565

PRIOR FILING DATE: 2001-01-08

PRIOR APPLICATION NUMBER: US 09/625,191

PRIOR FILING DATE: 2000-07-21

PRIOR APPLICATION NUMBER: US 09/543,371

PRIOR FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: US 09/479,118

PRIOR FILING DATE: 2000-01-07

PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/126,175

PRIOR FILING DATE: 1999-03-25

PRIOR APPLICATION NUMBER: US 60/089,689

PRIOR FILING DATE: 1998-06-17

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 22

LENGTH: 191

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)

US-10-032-221B-22

Query Match 100.0%; Score 147; DB 14; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.6e-13; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPLFCVNDVCFASRNDYSYWL 25

Db 20 TMAPLFCVNDVCFASRNDYSYWL 44

RESULT 8

US-10-270-877-46

Sequence 46, Application US/10270877

Publication No. US20030049791A1

GENERAL INFORMATION:

APPLICANT: Saus, Juan

TITLE OF INVENTION: Goodpasture Binding Protein

FILE REFERENCE: 98-723-AD1

CURRENT APPLICATION NUMBER: US/10/270,877

CURRENT FILING DATE: 2002-10-11

PRIOR APPLICATION NUMBER: 09/512,563

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/121,483

PRIOR FILING DATE: 1999-02-24

NUMBER OF SEQ ID NOS: 63

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 46

LENGTH: 211

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GPDV

US-10-270-877-46

Query Match 100.0%; Score 147; DB 14; Length 211;

Best Local Similarity 100.0%; Pred. No. 1.8e-13; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFAASNDYSYWL 25  
Db 73 TMAPFCNVNDVNCNFAASNDYSYWL 97

## RESULT 9

US-10-270-837-46  
; Sequence 46, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-837-46

Query Match 100.0%; Score 147; DB 14; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.9e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFAASNDYSYWL 25  
Db 73 TMAPFCNVNDVNCNFAASNDYSYWL 97

## RESULT 10

US-10-206-699-304  
; Sequence 304, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 304  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: alpha 3 chain  
US-10-206-699-304

Query Match 100.0%; Score 147; DB 14; Length 232;  
Best Local Similarity 100.0%; Pred. No. 1.9e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFAASNDYSYWL 25  
Db 61 TMAPFCNVNDVNCNFAASNDYSYWL 85

## RESULT 11

US-10-032-221B-10  
; Sequence 10, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-10

Query Match 100.0%; Score 147; DB 14; Length 244;  
Best Local Similarity 100.0%; Pred. No. 2e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNVNDVNCNFAASNDYSYWL 25  
Db 73 TMAPFCNVNDVNCNFAASNDYSYWL 97

## RESULT 12

US-10-206-699-302  
; Sequence 302, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 302  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
US-10-206-699-302

```
;
; OTHER INFORMATION: alpha 1 chain
; US-10-032-221B-2
;
; Query Match          95.9%; Score 141; DB 14; Length 229;
; Best Local Similarity 92.0%; Pred. No. 1.4e-12;
; Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
;      |||||:|||||
; Db 59 TMPFLFCNVNDVCFASRNDYSYWL 83
;
; RESULT 13
; US-10-032-221B-2
; ; Sequence 2, Application US/10032221B
; ; Publication No. US20030144481A1
; ; GENERAL INFORMATION:
; ; APPLICANT: Kalluri, Raghuram
; ; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; ; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; ; CURRENT APPLICATION NUMBER: US/10/032,221B
; ; PRIOR FILING DATE: 2001-12-21
; ; PRIOR APPLICATION NUMBER: PCT/US01/00565
; ; PRIOR FILING DATE: 2001-01-08
; ; PRIOR APPLICATION NUMBER: US 09/625,191
; ; PRIOR FILING DATE: 2000-07-21
; ; PRIOR APPLICATION NUMBER: US 09/543,371
; ; PRIOR FILING DATE: 2000-04-04
; ; PRIOR APPLICATION NUMBER: US 09/479,118
; ; PRIOR FILING DATE: 2000-01-07
; ; PRIOR APPLICATION NUMBER: US 09/335,224
; ; PRIOR FILING DATE: 1999-06-17
; ; PRIOR APPLICATION NUMBER: US 60/126,175
; ; PRIOR FILING DATE: 1999-03-25
; ; PRIOR APPLICATION NUMBER: US 60/089,689
; ; PRIOR FILING DATE: 1998-06-17
; ; NUMBER OF SEQ ID NOS: 58
; ; SOFTWARE: PatentIn version 3.1
; ; SEQ ID NO 2
; ; LENGTH: 229
; ; TYPE: PRT
; ; ORGANISM: Homo sapiens
; US-10-032-221B-2
;
; Query Match          95.9%; Score 141; DB 14; Length 229;
; Best Local Similarity 92.0%; Pred. No. 1.4e-12;
; Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
;      |||||:|||||
; Db 59 TMPFLFCNVNDVCFASRNDYSYWL 83
;
; RESULT 14
; US-09-925-302-507
; ; Sequence 507, Application US/09925302
; ; Patent No. US20020044941A1
; ; GENERAL INFORMATION:
; ; APPLICANT: Rosen et al.
; ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; ; FILE REFERENCE: PA104
; ; CURRENT APPLICATION NUMBER: US/09/925,302
; ; CURRENT FILING DATE: 2001-08-10
; ; PRIOR APPLICATION NUMBER: PCT/US00/05918
; ; PRIOR FILING DATE: 2000-03-08
; ; PRIOR APPLICATION NUMBER: 60/124,270
; ; PRIOR FILING DATE: 1999-03-12
; ; NUMBER OF SEQ ID NOS: 896
; ; SOFTWARE: PatentIn Ver. 2.0
; ; SEQ ID NO 507
; ; LENGTH: 406
; ; TYPE: PRT
; ; ORGANISM: Homo sapiens
```

```
;
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-302-507
;
; Query Match          95.9%; Score 141; DB 9; Length 406;
; Best Local Similarity 92.0%; Pred. No. 2.5e-12;
; Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
;      |||||:|||||
; Db 236 TMPFLFCNVNDVCFASRNDYSYWL 260
;
; RESULT 15
; US-10-372-683-8
; ; Sequence 8, Application US/10372683
; ; Publication No. US20040009171A1
; ; GENERAL INFORMATION:
; ; APPLICANT: GERRITSEN, MARY E.
; ; APPLICANT: PEALE JR., FRANKLIN V.
; ; APPLICANT: WU, THOMAS D.
; ; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; ; FILE REFERENCE: P1928R1P1
; ; CURRENT APPLICATION NUMBER: US/10/372,683
; ; CURRENT FILING DATE: 2003-02-21
; ; PRIOR APPLICATION NUMBER: US 10/271,690
; ; PRIOR FILING DATE: 2002-10-16
; ; PRIOR APPLICATION NUMBER: US 60/344,534
; ; PRIOR FILING DATE: 2001-10-18
; ; NUMBER OF SEQ ID NOS: 49
; ; SEQ ID NO 8
; ; LENGTH: 1669
; ; TYPE: PRT
; ; ORGANISM: Homo sapien
; US-10-372-683-8
;
; Query Match          95.9%; Score 141; DB 15; Length 1669;
; Best Local Similarity 92.0%; Pred. No. 1e-11;
; Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 TMPFLFCNVNDVCFASRNDYSYWL 25
;      |||||:|||||
; Db 1499 TMPFLFCNVNDVCFASRNDYSYWL 1523
;
; Search completed: April 5, 2004, 07:36:06
; Job time : 15.678 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 5.87167 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221B-37  
Perfect score: 147  
Sequence: 1 TNPFLFCNVNDVCFASNDYSYWL 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/6C COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	100.0	211	4	US-09-512-563C-46
2	147	100.0	218	2	US-08-399-889-25
3	147	100.0	218	3	US-09-187-364-25
4	147	100.0	218	3	US-09-439-897-4
5	147	100.0	268	4	US-09-589-927-6
6	147	100.0	268	4	US-09-277-665-6
7	147	100.0	268	4	US-09-589-987-6
8	146	99.3	471	2	US-08-399-889-24
9	146	99.3	471	3	US-09-187-364-24
10	146	99.3	471	3	US-09-439-897-2
11	141	95.9	260	4	US-09-589-927-2
12	141	95.9	260	4	US-09-277-665-2
13	141	95.9	260	4	US-09-589-987-2
14	139	94.6	264	4	US-09-589-927-10
15	139	94.6	264	4	US-09-277-665-10
16	139	94.6	264	4	US-09-589-987-10
17	115	78.2	260	4	US-09-589-927-12
18	115	78.2	260	4	US-09-277-665-12
19	115	78.2	260	4	US-09-589-987-12
20	112	76.2	258	4	US-09-589-927-4
21	112	76.2	258	4	US-09-277-665-4
22	112	76.2	258	4	US-09-589-987-4
23	103	70.1	260	4	US-09-589-927-8
24	103	70.1	260	4	US-09-277-665-8
25	103	70.1	260	4	US-09-589-987-8
26	90	61.2	1694	1	US-08-494-168-2
27	50.5	34.4	663	4	US-09-194-468A-30

28 47 32.0 186 2 US-08-766-551-3 Sequence 3, Appli  
29 46 31.3 400 4 US-03-134-000C-4389 Sequence 489, Ap  
30 45 30.6 410 4 US-09-543-861A-5407 Sequence 5407, Ap  
31 44.5 30.3 704 3 US-08-792-832A-2 Sequence 2, Appli  
32 44 29.9 49 1 US-07-865-166A-6 Sequence 6, Appli  
33 44 29.9 313 4 US-09-489-039A-8623 Sequence 8623, Ap  
34 44 29.9 525 4 US-09-549-519-26 Sequence 26, Appli  
35 43.5 29.6 164 4 US-09-634-338-381 Sequence 381, App  
36 43 29.3 103 1 US-08-271-562-1 Sequence 1, Appli  
37 43 29.3 103 1 US-08-087-007-3 Sequence 3, Appli  
38 43 29.3 103 2 US-08-696-777-1 Sequence 1, Appli  
39 43 29.3 103 3 US-08-483-433-3 Sequence 3, Appli  
40 43 29.3 103 5 PCT-US92-05920-3 Sequence 3, Appli  
41 43 29.3 105 3 US-09-591-435-12 Sequence 12, Appli  
42 43 29.3 128 6 5179198-1 Patent No. 5179198  
43 43 29.3 128 6 5521296-1 Patent No. 5521296  
44 43 29.3 160 4 US-09-668-673B-7 Sequence 7, Appli  
45 43 29.3 343 2 US-08-933-750C-13 Sequence 13, Appli

#### ALIGNMENTS

RESULT 1  
US-09-512-563C-46  
; Sequence 46, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512.563C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-09-512-563C-46

Query Match 100.0%; Score 147; DB 4; Length 211;  
Best Local Similarity 100.0%; Pred. No. 9.5e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNPFLFCNVNDVCFASNDYSYWL 25  
DB 73 TNPFLFCNVNDVCFASNDYSYWL 97

RESULT 2  
US-08-399-889-25  
; Sequence 25, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 218  
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match      100.0%; Score 147; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      100.0%; Score 147; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      100.0%; Score 147; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.8e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 47 TMPFLFCNVNDVCFASRNDYSYWL 71

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525I
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match      100.0%; Score 147; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 97 TMPFLFCNVNDVCFASRNDYSYWL 121

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-I
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match      100.0%; Score 147; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 97 TMPFLFCNVNDVCFASRNDYSYWL 121

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525I
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match      100.0%; Score 147; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFLFCNVNDVCFASRNDYSYWL 25
Db 97 TMPFLFCNVNDVCFASRNDYSYWL 121
```

```
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||
Db 97 TMAPFCNVNDVCFASRNDYSYWL 121

RESULT 8
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 99.3%; Score 146; DB 2; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 9
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match 99.3%; Score 146; DB 3; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 10
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match 99.3%; Score 146; DB 3; Length 471;
Best Local Similarity 96.0%; Pred. No. 3.1e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||
Db 300 TMAPFCNVNDVCFASRNDYSYWL 324

RESULT 11
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 95.9%; Score 141; DB 4; Length 260;
Best Local Similarity 92.0%; Pred. No. 8.7e-13;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMAPFCNVNDVCFASRNDYSYWL 25
* |||||:|||||
Db 90 TMAPFCNVNDVCFASRNDYSYWL 114

RESULT 12
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-2

Query Match 95.9%; Score 141; DB 4; Length 260;
Best Local Similarity 92.0%; Pred. No. 8.7e-13;
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.14528 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221b-38

Perfect score: 148

Sequence: 1 TMPFMFCINNVNCFASRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	148	100.0	253	2 I48304	collagen alpha 5(I)
2	148	100.0	754	2 A55267	collagen alpha 5(I)
3	148	100.0	1691	1 S22917	collagen alpha 5(I)
4	145	98.0	161	2 S49488	collagen alpha 3(I)
5	145	98.0	246	2 I48302	collagen alpha 3(I)
6	145	98.0	258	2 B61228	collagen alpha 1(I)
7	145	98.0	1689	1 CGHU4B	collagen alpha 1(I)
8	145	98.0	1689	1 CGMS4B	collagen alpha 1(I)
9	140	94.6	471	2 A39024	collagen alpha 3(I)
10	139	93.9	220	2 B49736	collagen alpha 3(I)
11	139	93.9	1670	1 CGHU3B	collagen alpha 3(I)
12	133	89.9	1752	2 A45407	collagen alpha 3(I)
13	129	87.2	1777	2 A54121	collagen alpha 3(I)
14	127	85.8	1783	2 S16366	collagen alpha-4 c
15	124	83.8	1758	2 T29350	hypothetical prote
16	124	83.8	1759	2 T29351	collagen alpha 2(I)
17	121	81.8	1744	2 S40991	collagen alpha 1(I)
18	120	81.1	261	2 A34476	collagen alpha 2(I)
19	113	76.4	1691	1 CGHU6B	collagen alpha 6(I)
20	104	70.3	312	2 I48303	collagen alpha 4(I)
21	104	70.3	623	2 A45137	collagen alpha 4(I)
22	104	70.3	775	2 A61228	collagen alpha 2(I)
23	104	70.3	1690	1 CGHU1B	collagen alpha 4(I)
24	104	70.3	1707	2 A33526	collagen alpha 2(I)
25	104	70.3	1712	1 CGHU2B	collagen alpha 2(I)
26	103	69.6	453	2 S18804	collagen alpha 4(I)
27	96	64.9	1761	2 T13990	collagen type IV a
28	95	64.2	1775	2 A31893	collagen alpha 1(I)
29	67	45.3	79	2 C43928	probable collagen

30 58.5 39.5 58 2 B43928 probable collagen  
31 50.5 34.1 257 2 H75419 hypothetical prote  
32 50 235 2 T23439 hypothetical prote  
33 49.5 33.4 886 2 T39081 hypothetical prote  
34 48.5 32.8 388 2 T29364 hypothetical prote  
35 48 32.4 523 2 T28727 hypothetical prote  
36 48 32.4 531 2 T28222 hypothetical prote  
37 48 32.4 1060 2 T30823 bumetanide sensiti  
38 47.5 32.1 261 2 D96772 probable RING zinc  
39 47.5 32.1 1743 2 T18279 multidrug resistan  
40 47 31.8 249 2 T04939 hypothetical prote  
41 47 31.8 356 2 T10514 probable stem brom  
42 47 31.8 366 2 S45915 hypothetical prote  
43 46.5 31.4 4981 2 T18489 hypothetical prote  
44 46 31.1 98 1 F2NTK photoystem II pro  
45 46 31.1 118 2 T37269 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

I48304

collagen alpha 5(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999

C:Accession: I48304; S47280

R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A:Molecule type: mRNA

A:Status: preliminary

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 100.0%; Score 148; DB 2; Length 253;

Best Local Similarity 100.0%; Pred. No. 2,3e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25

DB 83 TMPFMFCINNVNCFASRNDYSYWL 107

##### RESULT 2

A55267

collagen alpha 5(IV) chain - dog (fragment)

C:Species: Canis lupus familiaris (dog)

C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999

C:Accession: A55267

R:Zheng, K.; Thorner, P.S.; Marrano, P.; Baumal, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-

en type IV.

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>

A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548

C:Superfamily: collagen alpha 1(IV) chain

Query Match 100.0%; Score 148; DB 2; Length 754;

Best Local Similarity 100.0%; Pred. No. 6,3e-13;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25

DB 591 TMPFMFCINNVNCFASRNDYSYWL 615

## RESULT 3

S22917 collagen alpha 5(IV) chain precursor, renal splice form - human  
N/Alternate names: procollagen alpha 5(IV) chain  
N/Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C/Species: Homo sapiens (man)  
C/Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 21-Jul-2000  
C/Accession: S22917, A54365, A57079; A37122, 154317; A34850; S18850; I56971, 176598; A357079, J. Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A/Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identification of a novel splice form  
A/Reference number: S22917; MUID:92316923; PMID:1352287  
A/Accession: S22917  
A/Molecule type: mRNA  
A/Residues: 1-967 <ZH0>  
A/Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234  
R/Zhou, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A/Title: Structure of the human type IV collagen COL4A5 gene.  
A/Reference number: A54365; MUID:94165049; PMID:8120014  
A/Accession: A54365  
A/Molecule type: DNA  
A/Residues: 1-922 <ZH2>  
A/Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; PID:G587204  
R/Zhou, J.; Mochizuki, T.; Pihlajaniemi, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggvason, K.  
Science 261, 1167-1169, 1993  
A/Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited scleroderma  
A/Reference number: A57079; MUID:93361972; PMID:8356449  
A/Accession: A57079  
A/Molecule type: DNA  
A/Residues: 1-27 <ZH4>  
A/Cross-references: GB:Z37153; NID:G587203; PIDN:CAA85512.1; PID:G587204  
R/Pihlajaniemi, T.; Pihlajaniemi, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A/Title: Complete primary structure of the triple-helical region and the carboxyl-terminal region of the COL4A5 gene  
A/Reference number: A37122; MUID:90337990; PMID:2380186  
A/Accession: A37122  
A/Molecule type: mRNA  
A/Residues: 84-439, 'GS', 442-624, 'LALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>  
A/Cross-references: GB:J05558; EMBL:M58526; NID:G1314209  
A/Note: Submitted to the EMBL Data Library, February 1991  
A/Note: The authors translated the codon GCC for residue 115 as Val  
R/Renier, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Maessene, J.  
Hum. Mol. Genet. 1, 127-129, 1992  
A/Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in a patient with Alport syndrome  
A/Reference number: I54317; MUID:93244772; PMID:1363780  
A/Accession: I54317  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 313-324, 'E', 326-330 <REN>  
A/Cross-references: GB:559334; NID:G929946; PIDN:AAD13909.1; PID:G4261609  
R/Hosikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
A/Title: Identification of a distinct type IV collagen alpha chain with restricted kidney expression  
A/Reference number: A34850; MUID:90160375; PMID:1689491  
A/Accession: A34850  
A/Molecule type: mRNA  
A/Residues: 914-1264, 1271-1691 <HOS>  
A/Cross-references: EMBL:M31115; NID:G180824; PIDN:AAA52045.1; PID:G180825  
R/Zhou, J.; Hosikka, S.L.; Chow, L.T.; Tryggvason, K.  
Genomics 9, 1-9, 1991  
A/Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that encodes the alpha-5(IV) chain  
A/Reference number: A37969; MUID:91169491; PMID:2004755  
A/Accession: S18850  
A/Molecule type: DNA  
A/Residues: 924-1264, 1271-1691 <ZH3>  
A/Cross-references: EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:G177922; PIDN:AAA51858.1; PID:G177922  
R/Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; Mochizuki, T.  
Kidney Int. 44, 1316-1321, 1993  
A/Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex

A/Reference number: I56971; MUID:94133540; PMID:8301933  
A/Accession: I56971  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1258-1276 <GUOI>  
A/Cross-references: GB:569168; NID:G545095; PIDN:AAC60612.1; PID:G545096  
A/Note: kidney splice form  
A/Accession: I76598  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1284-1291, 'TFLGYLACLV' <GUO2>  
A/Cross-references: GB:569169; NID:G545097; PIDN:AAC60613.1; PID:G545098  
A/Note: frameshift mutation in patient with Alport syndrome  
R/Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; Saito, T.; Hum, Genet. 45, 1024-1033, 1990  
A/Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the region 10p11-10p12  
A/Reference number: A35335; MUID:90252791; PMID:2339699  
A/Accession: A35335  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1448-1477 <MYE>  
R/Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Y. Kidney Int. 46, 1307-1314, 1994  
A/Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primor  
A/Reference number: I56975; MUID:95156893; PMID:7853788  
A/Accession: I56975  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1595-1602 <NAK>  
A/Cross-references: GB:575903; NID:G913882; PIDN:AAB33374.1; PID:G913883  
A/Note: premature termination mutation from a patient with Alport syndrome; one other t  
R/Leimink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.  
Genomics 17, 485-489, 1993  
A/Title: Identification of four novel mutations in the COL4A5 gene of patients with Al  
A/Reference number: I54188; MUID:94010948; PMID:8406498  
A/Accession: I54188  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1604-1607, 'VRDAYKC' <LEM>  
A/Cross-references: GB:565767; NID:G425563; PIDN:AAD13967.1; PID:G4261667  
A/Note: frameshift mutation from a patient with Alport syndrome; five other mutations  
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
and subsequently O-glycosylated.  
C/Genetics:  
A/Gene: GDB:COL4A5; ATS  
A/Cross-references: GDB:120596; OMIM:303630  
A/Map position: Xq22-Xq22  
A/Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 225/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1;  
A/Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands wi  
C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-  
C/Function:  
A/Description: minor structural component of extracellular basement membrane  
C/Superfamily: collagen alpha 1(IV) chain  
C/Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; gly  
F1-26/Domain: signal sequence #status predicted <SIG>  
F127-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MA  
F127-1264, 1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status  
F127-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status  
F125/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F1592-1570, 1515-1573/Disulfide bonds: (or 1489-1573, 1515-1570) #status predicted  
F1592-1570, 1515-1573/Disulfide bonds: (or 1489-1573, 1515-1570) #status predicted  
F1592-1570, 1515-1573/Disulfide bonds: (or 1489-1573, 1515-1570) #status predicted  
F1592-1684, 1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted  
Query Match 100.0%; Score 148; DB 1; Length 1691;  
Best Local Similarity 100.0%; Pred. No. 1.3e-12;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASRNDYSYL 25

Db 1521 TMAPFCNNVNCNFCASRNDYSYL 1545

## RESULT 4

S49488

collagen alpha 3(IV) chain - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999

C:Accession: S49488

R:Oberbaumer, I.

submitted to the EMBL Data Library, October 1994

A:Description: Cloning of the NCI domains to the minor collagen IV chains of mouse via F cells.

A:Reference number: S49487

A:Accession: S49488

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-161 &lt;OB&gt;

A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916

C:Superfamily: collagen alpha 1(IV) chain

Query Match 98.0%; Score 145; DB 2; Length 161;

Best Local Similarity 96.0%; Pred. No. 3.9e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASRNDYSYL 25

Db 8 TMAPFCNNVNCNFCASRNDYSYL 32

## RESULT 5

I48302

collagen alpha 3(IV) chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997

C:Accession: I48302; S47278

R:Miner, J.H.; Sane, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence and organization of the alpha 3 chain

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48302

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-246 &lt;RES&gt;

A:Cross-references: EMBL:X35186; NID:G535187; PID:G535198

C:Superfamily: collagen alpha 1(IV) chain

Query Match 98.0%; Score 145; DB 2; Length 246;

Best Local Similarity 96.0%; Pred. No. 5.8e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASRNDYSYL 25

Db 75 TMAPFCNNVNCNFCASRNDYSYL 99

## RESULT 6

B61228

collagen alpha 1(IV) chain - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 17-Mar-1999

C:Accession: B61228

R:Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.

Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991

A:Title: Cloning of alpha(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelial cells

A:Reference number: A61228; MUID:92010685; PMID:1717398

A:Accession: B61228

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-258 <YAM>  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 98.0%; Score 145; DB 2; Length 258;

Best Local Similarity 96.0%; Pred. No. 6.1e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASRNDYSYL 25

Db 88 TMAPFCNNVNCNFCASRNDYSYL 112

## RESULT 7

CGH48

collagen alpha 1(IV) chain precursor - human

A:Alternate names: procollagen alpha 1(IV) chain

C:Species: Homo sapiens (man)

C:Date: 28-May-1986 #sequence\_revision 31-Dec-1992 #text\_change 07-Dec-1999

C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S0207; S39614; A02863; A02863; K.

R:Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 264, 13565-13571, 1989

A:Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen

A:Reference number: S16876; MUID:89340433; PMID:2701944

A:Accession: S16876

A&gt;Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1669 &lt;SOI1&gt;

A:Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAAS3098.1; PID:G180803

A&gt;Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988

R:Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen

A:Reference number: A92690; MUID:89034231; PMID:3182844

A:Accession: A32117

A:Molecule type: DNA

A:Residues: 1-28 &lt;SOI2&gt;

A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G553233

R:Poschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane

A:Reference number: S02738; MUID:89030632; PMID:2846280

A:Accession: S02738

A&gt;Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-6, 'L', 8-28 &lt;PO&gt;

A:Cross-references: EMBL:X12784; NID:G30072

R:Brazel, D.; Oberbaumer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.

Eur. J. Biochem. 168, 529-536, 1987

A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement membrane

A:Reference number: S00048; MUID:88029471; PMID:3311751

A:Accession: S00048

A:Molecule type: mRNA

A:Residues: 1-318, 'A', 320-944 &lt;BRAL&gt;

A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067

A:Accession: S25826

A:Molecule type: protein

A:Residues: 271-318, 'A', 320-554 &lt;BRAL&gt;

R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (75-100) of human type IV collagen

A:Reference number: A23115; MUID:86004708; PMID:4043082

A:Accession: A23115

A:Molecule type: protein

A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 &lt;GLA&gt;

A:Experimental source: placenta

A&gt;Note: the amino end of the mature form is blocked

R:Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.

FEBS Lett. 225, 188-194, 1987

A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane

A:Reference number: S00207; MUID:88083584; PMID:3691802

A:Accession: S00207

A:Molecule type: mRNA

A:Residues: 244-530 &lt;SOI3&gt;



A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
 R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
 EMO J. 12, 4795-4802, 1993  
 A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen  
 A;Reference number: S39614; MUID:94038963; PMID:8223488  
 A;Accession: S39614  
 A;Molecule type: protein  
 A;Residues: 371-554 <EBL>  
 R;Babel, W.; Glanville, R.W.  
 Eur. J. Biochem. 143, 545-556, 1984  
 A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se  
 A;Reference number: A02863; MUID:85003629; PMID:6434307  
 A;Accession: A02863  
 A;Molecule type: protein  
 A;Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-  
 A;Experimental source: placenta  
 R;Glanville, R.W.; Rauter, A.  
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
 A;Title: Peptin fragments of human placental basement-membrane collagens showing interru  
 A;Reference number: S16908; MUID:82005835; PMID:6792033  
 A;Accession: A58517  
 A;Molecule type: protein  
 A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14  
 R;MacWhirt, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
 Biochemistry 22, 4940-4948, 1983  
 A;Title: Isolation and characterization of pepsin-solubilized human basement membrane (b  
 A;Reference number: S16910; MUID:84053346; PMID:6416291  
 A;Accession: S16910  
 A;Molecule type: protein  
 A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-  
 A;Experimental source: placenta  
 R;Philajantemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F  
 J. Biol. Chem. 260, 7681-7687, 1985  
 A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen  
 A;Reference number: S01466; MUID:85207819; PMID:2581569  
 A;Accession: S01466  
 A;Molecule type: mRNA  
 A;Residues: 1256-1669 <PIH>  
 A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424  
 R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
 A;Title: Restricted homology between human alpha-1 type IV and other procollagen chains.  
 A;Reference number: S16879; MUID:85216555; PMID:2582422  
 A;Accession: S16879  
 A;Molecule type: mRNA  
 A;Residues: 1259-1669 <BRI>  
 R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,  
 Eur. J. Biochem. 147, 217-224, 1985  
 A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
 A;Reference number: A02864; MUID:85127033; PMID:2578961  
 A;Accession: S19091  
 A;Molecule type: protein  
 A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',  
 R;Siebold, B.; Deutzmann, R.; Kuehn, K.  
 Eur. J. Biochem. 176, 617-624, 1988  
 A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxy-term  
 A;Reference number: S02850; MUID:89005112; PMID:2844531  
 A;Contents: annotation; disulfide bonds  
 C;Genetics:  
 A;Gene: GDB:COL4A1  
 A;Cross-references: GDB:119791; OMIM:120130  
 A;Map position: 13q34-13q34  
 A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/  
 1/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116  
 C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(  
 C;Associations among trimer amino-terminal domains (disulfide and desmosine cross-links), dim  
 C;Associations in the interrupted helical domain (with disulfide and desmosine cr  
 C;Function:  
 A;Description: structural component of extracellular basement membrane  
 C;Superfamily: collagen alpha 1(IV) chain  
 C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication  
 F;1-26/Domain: signal sequence #status predicted <SIG>

F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <WAT>  
 F;29-162/Domain: amino-terminal nonhelical, 7S <YSD>  
 F;163-1440/Domain: interrupted helical <COL>  
 F;414-452/Region: integrin binding #status experimental  
 F;597-599/Region: cell attachment (R-G-D) motif  
 F;917-919/Region: cell attachment (R-G-D) motif  
 F;968-970/Region: cell attachment (R-G-D) motif  
 F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NCL>  
 F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTL>  
 F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>  
 F;27/Modified site: blocked amino end (Ala) (in mature form) #status experimental  
 F;31,36,39,41,125,434,167,470/Disulfide bonds: interchain #status predicted  
 F;45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543,5  
 1081,1084,1099,1117,1132,1150,1165,1182,1185,1188,1226,1235,1265,1283,1304,1319,1328,13  
 F;45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582,6  
 99,1117,1132,1150,1165,1182,1185,1206,1235,1265,1283,1304,1319,1328,1340,1356,1371  
 F;54,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,184  
 9,745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F;126/Binding site: carboxylate (Asn) (covalent) #status experimental  
 F;129/Modified site: allysine (Lys) #status predicted  
 F;172,540,547/Modified site: 5-hydroxylysine (Lys) #status atypical  
 F;272,635,839/Modified site: 4-hydroxyproline (Pro) #status atypical  
 F;446-447/Cleavage site: Gly-Ile (Gelatinase B) #status predicted  
 F;766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899,9  
 23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256  
 431,1429/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F;1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F;1120,1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental  
 F;1120,1424/Modified site: 3-hydroxyproline (Pro) #status absent  
 F;1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F;1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F;1505-1511,1616-1622/Disulfide bonds: #status predicted  
 F;1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
 Query Match 98.0%; Score 145; DB 1; Length 1669;  
 Best Local Similarity 96.0%; Pred. No. 3,5e-12;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25  
 DB 1499 TMPFLFCNNVNCVFASRNDYSYWL 1523  
 RESULT 8  
 CGMS4B  
 Collagen alpha 1(IV) chain precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000  
 F;1460-1548,1493-1551/Disulfide bonds: A02864; A25636; A29301; S19079; A32003; A31766; S1  
 R;Muthukumar, G.; Blumberg, B.; Kurkinen, M.  
 J. Biol. Chem. 264, 6310-6317, 1989  
 A;Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Dif  
 A;Reference number: A33525; MUID:89197932; PMID:2703490  
 A;Accession: A33525  
 A;Molecule type: mRNA  
 A;Residues: 1-1669 <MUT>  
 A;Cross-references: EMBL:J04694; NID:G556296; PIDN:AAAS0292.1; PID:G556297  
 R;Wood, L.; Theriault, N.; Vogeli, G.  
 FEBS Lett. 227, 5-8, 1988  
 A;Title: cDNA clones completing the nucleotide and derived amino acid sequence of the a  
 A;Reference number: S01454; MUID:88112221; PMID:3338568  
 A;Accession: S01454  
 A;Molecule type: mRNA  
 A;Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-7  
 A;Cross-references: EMBL:X06777  
 R;Killen, P.D.; Buelo, P.; Sakurai, Y.; Yamada, Y.  
 J. Biol. Chem. 263, 8706-8709, 1988  
 A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha  
 A;Reference number: A28066; MUID:88243724; PMID:3379041  
 A;Accession: A28066  
 A;Molecule type: mRNA  
 A;Residues: 1-129 <KIT>



Best Local Similarity 88.0%; Pred. No. 3.6e-12;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMFMFCNNVNCVFASNDYSYWL 25  
|||||:|||||  
DB 82 TMFFLCNVNDCVFASNDYSYWL 106  
|||||:|||||

RESULT 11  
CGHU3B  
collagen alpha 3 (IV) chain precursor, long splice form - human  
N;Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
C;Species: Homo sapiens (man)  
C;Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
C;Accession: A54763; A43928; A44043; A45971; A39786  
R;Moriyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Readler, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
A;Reference number: A54763; MUID:94364994; PMID:8083201  
A;Accession: A54763  
A;Molecule type: mRNA  
A;Residues: 1-1670 <MAR>  
A;Cross-references: GB:X80031; NID:G577563; PID:G577564  
A;Experimental source: kidney  
R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the  
A;Reference number: A43928; MUID:92147878; PMID:1737849  
A;Accession: A43928  
A;Molecule type: mRNA  
A;Residues: 1331-1524, 'I', 1526-1670 <TUR>  
A;Cross-references: GB:M81379  
A;Experimental source: kidney  
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 15780-15784, 1992  
A;Title: Exon/intron structure of the human alpha 3 (IV) gene encompassing the Goodpasture  
A;Reference number: A44043; MUID:93015926; PMID:1400291  
A;Accession: A44043  
A;Molecule type: DNA; mRNA  
A;Residues: 1386-1670 <QUI>  
A;Cross-references: GB:M92993; NID:G177895; PIDN:AA21610.1; PID:G177896  
A;Note: sequence extracted from NCBI backbone (NCBIP:115597)  
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 269, 17358, 1994  
A;Reference number: A44738; MUID:94274734; PMID:8006044  
A;Contents: annotation; erratum; correction to intronic sequence in A44043  
R;Bernal, D.; Quinones, S.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A;Reference number: A45971; MUID:93280184; PMID:8505332  
A;Accession: A45971  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1427-1444 <BER>  
A;Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly iden  
R;Morrisson, K.E.; Maruyama, M.; Yang-Feng, T.L.; Readler, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1991  
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain o  
A;Reference number: A39786; MUID:91353570; PMID:1882840  
A;Accession: A39786  
A;Molecule type: mRNA  
A;Residues: 1453-1593, 'A', 1595-1670 <WOR>  
A;Cross-references: GB:S55790; NID:G234418; PIDN:AA19637.1; PID:G234419  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitop  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-q37  
A;Introns: 1385/1; 1418/1; 1489/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A;Note: the alpha 3 (IV) and alpha 4 (IV) chain genes are encoded on opposite strands w

A;introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with

F:1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

Query March 95 84. Score 127. DE 2. Length 1708  
F:1593-1599,1702-1708/disulfide bonds: #status predicted  
F:136/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91,93,95,97,99,101,103,105,107,109,111,113,115,117,119,121,123,125,127,129,131,133,135,137,139,141,143,145,147,149,151,153,155,157,159,161,163,165,167,169,171,173,175,177,179,181,183,185,187,189,191,193,195,197,199,201,203,205,207,209,211,213,215,217,219,221,223,225,227,229,231,233,235,237,239,241,243,245,247,249,251,253,255,257,259,261,263,265,267,269,271,273,275,277,279,281,283,285,287,289,291,293,295,297,299,301,303,305,307,309,311,313,315,317,319,321,323,325,327,329,331,333,335,337,339,341,343,345,347,349,351,353,355,357,359,361,363,365,367,369,371,373,375,377,379,381,383,385,387,389,391,393,395,397,399,401,403,405,407,409,411,413,415,417,419,421,423,425,427,429,431,433,435,437,439,441,443,445,447,449,451,453,455,457,459,461,463,465,467,469,471,473,475,477,479,481,483,485,487,489,491,493,495,497,499,501,503,505,507,509,511,513,515,517,519,521,523,525,527,529,531,533,535,537,539,541,543,545,547,549,551,553,555,557,559,561,563,565,567,569,571,573,575,577,579,581,583,585,587,589,591,593,595,597,599,601,603,605,607,609,611,613,615,617,619,621,623,625,627,629,631,633,635,637,639,641,643,645,647,649,651,653,655,657,659,661,663,665,667,669,671,673,675,677,679,681,683,685,687,689,691,693,695,697,699,701,703,705,707,709,711,713,715,717,719,721,723,725,727,729,731,733,735,737,739,741,743,745,747,749,751,753,755,757,759,761,763,765,767,769,771,773,775,777,779,781,783,785,787,789,791,793,795,797,799,801,803,805,807,809,811,813,815,817,819,821,823,825,827,829,831,833,835,837,839,841,843,845,847,849,851,853,855,857,859,861,863,865,867,869,871,873,875,877,879,881,883,885,887,889,891,893,895,897,899,901,903,905,907,909,911,913,915,917,919,921,923,925,927,929,931,933,935,937,939,941,943,945,947,949,951,953,955,957,959,961,963,965,967,969,971,973,975,977,979,981,983,985,987,989,991,993,995,997,999,1001,1003,1005,1007,1009,1011,1013,1015,1017,1019,1021,1023,1025,1027,1029,1031,1033,1035,1037,1039,1041,1043,1045,1047,1049,1051,1053,1055,1057,1059,1061,1063,1065,1067,1069,1071,1073,1075,1077,1079,1081,1083,1085,1087,1089,1091,1093,1095,1097,1099,1101,1103,1105,1107,1109,1111,1113,1115,1117,1119,1121,1123,1125,1127,1129,1131,1133,1135,1137,1139,1141,1143,1145,1147,1149,1151,1153,1155,1157,1159,1161,1163,1165,1167,1169,1171,1173,1175,1177,1179,1181,1183,1185,1187,1189,1191,1193,1195,1197,1199,1201,1203,1205,1207,1209,1211,1213,1215,1217,1219,1221,1223,1225,1227,1229,1231,1233,1235,1237,1239,1241,1243,1245,1247,1249,1251,1253,1255,1257,1259,1261,1263,1265,1267,1269,1271,1273,1275,1277,1279,1281,1283,1285,1287,1289,1291,1293,1295,1297,1299,1301,1303,1305,1307,1309,1311,1313,1315,1317,1319,1321,1323,1325,1327,1329,1331,1333,1335,1337,1339,1341,1343,1345,1347,1349,1351,1353,1355,1357,1359,1361,1363,1365,1367,1369,1371,1373,1375,1377,1379,1381,1383,1385,1387,1389,1391,1393,1395,1397,1399,1401,1403,1405,1407,1409,1411,1413,1415,1417,1419,1421,1423,1425,1427,1429,1431,1433,1435,1437,1439,1441,1443,1445,1447,1449,1451,1453,1455,1457,1459,1461,1463,1465,1467,1469,1471,1473,1475,1477,1479,1481,1483,1485,1487,1489,1491,1493,1495,1497,1499,1501,1503,1505,1507,1509,1511,1513,1515,1517,1519,1521,1523,1525,1527,1529,1531,1533,1535,1537,1539,1541,1543,1545,1547,1549,1551,1553,1555,1557,1559,1561,1563,1565,1567,1569,1571,1573,1575,1577,1579,1581,1583,1585,1587,1589,1591,1593,1595,1597,1599,1601,1603,1605,1607,1609,1611,1613,1615,1617,1619,1621,1623,1625,1627,1629,1631,1633,1635,1637,1639,1641,1643,1645,1647,1649,1651,1653,1655,1657,1659,1661,1663,1665,1667,1669,1671,1673,1675,1677,1679,1681,1683,1685,1687,1689,1691,1693,1695,1697,1699,1701,1703,1705,1707,1709,1711,1713,1715,1717,1719,1721,1723,1725,1727,1729,1731,1733,1735,1737,1739,1741,1743,1745,1747,1749,1751,1753,1755,1757,1759,1761,1763,1765,1767,1769,1771,1773,1775,1777,1779,1781,1783,1785,1787,1789,1791,1793,1795,1797,1799,1801,1803,1805,1807,1809,1811,1813,1815,1817,1819,1821,1823,1825,1827,1829,1831,1833,1835,1837,1839,1841,1843,1845,1847,1849,1851,1853,185

Query Match 85.8%; Score 127; DB 2; Length 1763;  
Best Local Similarity 80.0%; Pred. No. 1.2e-09;

Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSARNDSYWL 25  
|||||:|||||:|||||  
Db 1587 TMPFLCFDNNVNCNFSARNDSYWL 1611

## RESULT 15

T29350  
hypothetical protein F01G12.5a - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C/Accession: T29350  
R;Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A/Description: The sequence of C. elegans cosmid F01G12.  
A/Reference number: Z20611  
A/Accession: T29350  
A/Status: preliminary; translated from GH/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1758 <WUX>  
A/Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a  
A/Experimental source: strain Bristol N2; clone F01G12  
C/Genetics:  
A/Gene: CESP:F01G12.5a  
A/Map position: X  
A/Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3  
C/Superfamily: collagen alpha 1(IV) chain

Query Match 83.8%; Score 124; DB 2; Length 1758;

Best Local Similarity 80.0%; Pred. No. 3.le-09;

Matches 20; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSARNDSYWL 25  
|||||:|||||:|||||  
Db 1585 TMPFLCFDNNVNCNFSARNDSYWL 1609

Search completed: April 5, 2004, 07:05:36  
Job time : 6.14528 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.1477 Seconds

(without alignments)

413.557 Million cell updates/sec

Title: US-10-032-221b-38

Perfect score: 148

Sequence: 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	148	100.0	754	1 CA54 CANFA	Q28247 canis faml
2	148	100.0	1685	1 CA54 HUMAN	P29400 homo sapien
3	145	98.0	1669	1 CA14 HUMAN	P02462 mus sapien
4	145	98.0	1669	1 CA14 MOUSE	P02463 mus musculus
5	140	94.6	1471	1 CA34 BOVIN	Q01955 bos taurus
6	139	93.9	1671	1 CA34 HUMAN	Q01955 homo sapien
7	127	85.8	1763	1 CA24 ASCSU	P27393 ascaris suu
8	121	81.8	1758	1 CA14 CAEEL	P17139 caenorhabdi
9	120	81.1	1758	1 CA24 CAEEL	P17140 caenorhabdi
10	113	76.4	1691	1 CA64 HUMAN	Q14031 homo sapien
11	104	70.3	623	1 CA44 RABIT	P55787 oryctolagus
12	104	70.3	1690	1 CA44 HUMAN	P53420 homo sapien
13	104	70.3	1707	1 CA24 MOUSE	P08122 mus musculus
14	104	70.3	1712	1 CA24 HUMAN	P08572 homo sapien
15	103	69.6	453	1 CA14 BOVIN	Q29442 bos taurus
16	95	64.2	1775	1 CA14 DROME	P08120 drosophila
17	49	33.1	333	1 AMR1 HUMAN	Q9Y4X0 homo sapien
18	49	33.1	344	1 AMR1 MOUSE	Q9Jht5 mus musculus
19	48	32.4	1060	1 NKCL MANSE	Q25479 manduca sex
20	47.5	32.1	1743	1 TAGC DICDI	Q23868 dictyosteli
21	47	31.8	366	1 MDM2 YEAST	P38236 saccharomyc
22	47	31.8	976	1 HMDH GIBFU	Q12577 gibberella
23	45.5	30.7	704	1 OE66 NPVAC	Q00704 autographa
24	45	30.4	397	1 YMP7 YEAST	Q04359 saccharomyc
25	45	30.4	867	1 SUL1 COTCO	Q90xb6 coturnix co
26	45	30.4	870	1 SUL1 MOUSE	Q8K007 mus musculus
27	45	30.4	870	1 SUL1 RAT	Q8vi60 rattus norv
28	45	30.4	871	1 SUL1 HUMAN	Q8viu6 homo sapien
29	44.5	30.1	433	1 TC01 HUMAN	P20061 homo sapien
30	44.5	30.1	743	1 NU5C CARCG	Q9t156 carpenteria
31	44	29.7	296	1 SAPR STRPU	P11761 strongyloce
32	44	29.7	308	1 META SALTI	Q8ziw1 salmonella
33	44	29.7	308	1 META SALTY	P37413 salmonella

34 44 29.7 359 1 LLCR\_SUNY3  
35 44 29.7 2273 1 ABCR\_HUMAN  
36 43.5 29.4 301 1 Y664\_METJA  
37 43.5 29.4 318 1 RLA0\_MAIZE  
38 43.5 29.4 663 1 NM02\_CHICK  
39 43.5 29.4 1850 1 VIT2\_CHICK  
40 43 29.1 256 1 PRN3\_HUMAN  
41 43 29.1 319 1 YN97\_CAEEL  
42 43 29.1 395 1 NH10\_CAEEL  
43 43 29.1 456 1 YC13\_ASTIO  
44 43 29.1 464 1 SYE2\_COXBU  
45 43 29.1 484 1 C24B\_PIG

#### ALIGNMENTS

RESULT 1  
CA54 CANFA  
ID CA54 CANFA STANDARD; PRT; 754 AA.  
AC Q28247;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Collagen alpha 5(IV) chain (Fragment).  
GN COL4A5  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Samoyed; TISSUE=Kidney;  
RX MEDLINE=9224866; PubMed=5171024;  
RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;  
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for  
human X-linked hereditary nephritis resulting from a single base  
mutation in the gene encoding the alpha 5 chain of collagen type  
IV";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
alpha 6(IV), each of which can form a triple helix structure with  
2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the G-  
X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
CC -!- PM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of  
canine X-linked hereditary nephritis (HN), a disease similar to  
that in humans (also referred to as Alport syndrome) characterized  
by progressive renal failure and neurosensory deafness.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----

DR EMBL; U07888; AAB60258.1; --  
DR PIR; A55267; A55267.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 8.  
DR ProDom; PD000007; C1g\_helix; 1.  
DR ProDom; PD03923; Procollagen4; 1.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
FT NON\_TER 1 1  
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.  
FT DOMAIN 531 >754 NON-HELICAL REGION (NC1).  
FT DISULFID 552 643 OR 640 (BY SIMILARITY).  
FT DISULFID 585 640 OR 643 (BY SIMILARITY).  
FT DISULFID 597 603 BY SIMILARITY.  
FT DISULFID 662 ? OR 754 (BY SIMILARITY).  
FT DISULFID 696 754 BY SIMILARITY.  
FT DISULFID 708 714 BY SIMILARITY.  
FT NON\_TER 754 754  
SQ SEQUENCE 754 AA; 73537 MW; DSE321C287FA925B CRC64;  
  
Query Match 100.0%; Score 148; DB 1; Length 754;  
Best-Local Similarity 100.0%; Pred. No. 5.7e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TPFMFCNNVNCNPFASRNDYSYWL 25  
Db 591 TPFMFCNNVNCNPFASRNDYSYWL 615  
  
RESULT 2  
CA54 HUMAN STANDARD; PRT; 1685 AA.  
ID CA54\_HUMAN STANDARD; PRT; 1685 AA.  
AC P29400; Q16006; Q16126;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 5 (IV) chain precursor.  
GN COL4A5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94165049; PubMed=8120014;  
RA Zhou J., Leinonen A., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A5 gene.";  
RL J. Biol. Chem. 269:6608-6614(1994).  
RN [2]  
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
RC TISSUE=Kidney;  
RX MEDLINE=92316923; PubMed=1352287;  
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";  
RL J. Biol. Chem. 267:12475-12481(1992).  
RN [3]  
RP SEQUENCE OF 85-1685 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=90337990; PubMed=2380186;  
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;  
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5 (IV).";  
RL J. Biol. Chem. 265:13759-13766(1990).  
RN [4]  
RP SEQUENCE OF 924-1685 FROM N.A.  
  
RX MEDLINE=91169491; PubMed=2004755;  
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";  
RL Genomics 9:1-9(1991).  
RN [5]  
RP SEQUENCE OF 914-1685 FROM N.A.  
RX MEDLINE=90160375; PubMed=1689491;  
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;  
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RP SEQUENCE OF 1442-1471 FROM N.A.  
RX MEDLINE=90252791; PubMed=2339699;  
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Shear D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome RT locus.";  
RL Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RP SEQUENCE OF 1-20 FROM N.A.  
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).  
RX MEDLINE=94133540; PubMed=8301933;  
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient RT deletes the NC1 domain.";  
RL Kidney Int. 44:1316-1321(1993).  
RN [9]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
RL Hum. Mutat. 9:477-499(1997).  
RN [10]  
RP VARIANT AS SER-1564.  
RX MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";  
RL Genomics 9:10-18(1991).  
RN [11]  
RP VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [12]  
RP VARIANT AS GLU-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [13]  
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=8406498;  
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,



RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
 RA van Oost B.A., Smeets H.J.M.; Identification of four novel mutations in the COL4A5 gene of  
 RT patients with Alport syndrome.";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [14]  
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;  
 RP ARG-869; ARG-872 AND CYS-1241.  
 RX MEDLINE=95322976; PubMed=7596311;  
 RA Boye E., Flierer F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the  
 RL COL4A5 gene in Alport syndrome patients.";  
 RL Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=96213750; PubMed=8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 RA Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RT common in the western United States.";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
 RA Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RT exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
 RP MET-1428.  
 RX MEDLINE=97094179; PubMed=8940267;  
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jaccaesier D.,  
 RA Giarras I., Drouot L., Descheres G., Gruenfeld J.-P., Broyer M.,  
 RA Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RT syndrome.";  
 RL Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-1498.  
 RX MEDLINE=96233932; PubMed=8829632;  
 RA Tverskaya S., Bobryna V., Tealykova F., Ignatova M.,  
 RA Krasnopolskaya X., Evgrafov O.;  
 RT "Substitution of A1498D in noncollagen domain of a5(IV) collagen  
 RT chain associated with adult-onset X-linked Alport syndrome.";  
 RL Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-1677.  
 RX MEDLINE=97295089; PubMed=9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RT syndrome and COL4A5 R1677Q.";  
 RL Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
 RP AND ASP-1596.  
 RX MEDLINE=98112435; PubMed=9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
 RA Trivelli A., Pinciatoli A.R., Ragoiolo M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RT Alport syndrome.";  
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANTS AS VAL-420; 456-PRO-PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
 RP 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
 RP ARG-1156; GLU-1261; SER-1337 AND ARG-1649.  
 RX MEDLINE=99063529; PubMed=9848783;  
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,

RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RT Alport syndrome using PCR and direct DNA sequencing.";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
 RX MEDLINE=20030197; PubMed=10561141;  
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
 RT "Detection of mutations in the COL4A5 gene in over 90% of male  
 RT patients with X-linked Alport's syndrome by RT-PCR and direct  
 RT sequencing.";  
 RL Am. J. Kidney Dis. 34:854-862(1999).  
 RN [23]  
 RP VARIANT AS ARG-822.  
 RN  
 Query Match 100.0%; Score 148; DB 1; Length 1685;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-12;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TMPEFMCNINNVNCFASRNDYSYWL 25  
 DB 1515 TMPEFMCNINNVNCFASRNDYSYWL 1539  
 RESULT 3  
 CA14\_HUMAN STANDARD; PRT; 1669 AA.  
 ID CA14\_HUMAN  
 AC P02462;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1] SEQUENCE FROM N.A.  
 RP MEDLINE=69340433; PubMed=2701944;  
 RX Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;  
 RT "Structural organization of the gene for the alpha 1 chain of human  
 RT type IV collagen.";  
 RL J. Biol. Chem. 264:13565-13571(1989).  
 RN [2]  
 RP SEQUENCE OF 46-1257 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88083584; PubMed=3691802;  
 RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
 RT "Complete primary structure of the alpha 1-chain of human basement  
 RT membrane (type IV) collagen.";  
 RL FEBS Lett. 225:188-194(1987).  
 RN [3]  
 RP SEQUENCE OF 1-943 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88029471; PubMed=3311751;  
 RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,  
 RA Deutzmann R., Kuehn K.;  
 RT "Completion of the amino acid sequence of the alpha 1 chain of human  
 RT basement membrane collagen (type IV) reveals 21 non-triplet  
 RT interruptions located within the collagenous domain.";  
 RL Eur. J. Biochem. 168:529-536(1987).  
 RN [4]  
 RP SEQUENCE OF 28-243.  
 RC TISSUE=Placenta;  
 RX MEDLINE=8604708; PubMed=4043082;  
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
 RT region (7S domain) of the alpha 1 (IV) chain of human basement  
 RT membrane collagen.";  
 RL Eur. J. Biochem. 152:213-219(1985).  
 RN [5]

RP SEQUENCE OF 534-1447.  
RX MEDLINE=85003629; PubMed=6434307;  
RA Babel W., Glanville R.W.;  
RT "Structure of human-basement-membrane (type IV) collagen. Complete  
RT amino-acid sequence of a 914-residue-long pepsin fragment from the  
RL alpha 1(IV) chain.";  
RL Eur. J. Biochem. 143:545-556(1984).  
RN [6]  
RP SEQUENCE OF 1256-1669 FROM N.A.  
RX MEDLINE=85207819; PubMed=2581969;  
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
RT "cDNA clones coding for the pro-alpha 1(IV) chain of human type IV  
RT procollagen reveal an unusual homology of amino acid sequences in two  
RT halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687(1985).  
RN [7]  
RP SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RT Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
RT procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89034231; PubMed=3182844;  
RA Solinen R., Huotari M., Hestikka S.L., Prockop D.J., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
RT collagen are divergently encoded on opposite DNA strands and have an  
RT overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220(1988).  
RN [9]  
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta; PubMed=2844531;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RT of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624(1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M26576; AAA53098.1; JOINED.  
CC EMBL; J04217; AAA53098.1; JOINED.  
CC EMBL; M26550; AAA53098.1; JOINED.

DR EMBL; M26540; AAA53098.1; JOINED.  
DR EMBL; M26542; AAA53098.1; JOINED.  
DR EMBL; M26543; AAA53098.1; JOINED.  
DR EMBL; M26544; AAA53098.1; JOINED.  
DR EMBL; M26545; AAA53098.1; JOINED.  
DR EMBL; M26546; AAA53098.1; JOINED.  
DR EMBL; M26547; AAA53098.1; JOINED.  
DR EMBL; M26537; AAA53098.1; JOINED.  
DR EMBL; M26538; AAA53098.1; JOINED.  
DR EMBL; M26548; AAA53098.1; JOINED.  
DR EMBL; M26549; AAA53098.1; JOINED.  
DR EMBL; M26551; AAA53098.1; JOINED.  
DR EMBL; M26552; AAA53098.1; JOINED.  
DR EMBL; M26553; AAA53098.1; JOINED.  
DR EMBL; M26554; AAA53098.1; JOINED.  
DR EMBL; M26555; AAA53098.1; JOINED.  
DR EMBL; M26556; AAA53098.1; JOINED.  
DR EMBL; M26557; AAA53098.1; JOINED.  
DR EMBL; M26558; AAA53098.1; JOINED.  
DR EMBL; M26559; AAA53098.1; JOINED.  
DR EMBL; M26560; AAA53098.1; JOINED.  
DR EMBL; M26561; AAA53098.1; JOINED.  
DR EMBL; M26562; AAA53098.1; JOINED.  
DR EMBL; M26563; AAA53098.1; JOINED.  
DR EMBL; M26564; AAA53098.1; JOINED.  
DR EMBL; M26565; AAA53098.1; JOINED.  
DR EMBL; M26566; AAA53098.1; JOINED.  
DR EMBL; M26567; AAA53098.1; JOINED.  
DR EMBL; M26568; AAA53098.1; JOINED.  
DR EMBL; M26569; AAA53098.1; JOINED.  
DR EMBL; M26570; AAA53098.1; JOINED.  
DR EMBL; M26571; AAA53098.1; JOINED.  
DR EMBL; M26572; AAA53098.1; JOINED.  
DR EMBL; M26573; AAA53098.1; JOINED.  
DR EMBL; M26574; AAA53098.1; JOINED.  
DR EMBL; M26575; AAA53098.1; JOINED.  
DR EMBL; Y00706; CAA68698.1; -  
DR EMBL; X05561; CAA29075.1; -  
DR EMBL; M10940; AAA52006.1; -  
DR EMBL; M11315; AAA52042.1; -  
DR PIR; S16876; CGHU4B.  
DR Genew; HGNC:2202; COL4A1.  
DR MTM; 120130; -  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4\_2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; Clg\_helix; 6.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Extracellular matrix; Connective tissue; Basement membrane;  
DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
DR SIGNAL 1 27 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
DR PROPEP 28 172 COLLAGEN ALPHA 1(IV) CHAIN.  
DR CHAIN 173 1669 TRIPLE-HELICAL REGION.  
DR DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
DR DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
DR DOMAIN 126 126 N-LINKED (GLCNAC. .).  
DR CARBOHYD 126 126 OR 1548.  
DR DISULFID 1460 1551 OR 1548.  
DR DISULFID 1493 1548 OR 1551.  
DR DISULFID 1505 1511 OR 1665.  
DR DISULFID 1570 1665 OR 1665.  
DR DISULFID 1604 1662 OR 1662.  
DR DISULFID 1616 1622 OR 1622.  
DR CONFLICT 237 238 SG -> KE (IN REF. 4).  
DR CONFLICT 241 241 G -> K (IN REF. 4).  
DR CONFLICT 319 319 Q -> A (IN REF. 3).  
DR CONFLICT 719 719 N -> D (IN REF. 5).  
DR CONFLICT 837 837 D -> Y (IN REF. 5).

```

FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BBE6DFFB9B9A84 CRC64;

Query Match 98.0%; Score 145; DB 1; Length 1669;
Best Local Similarity 96.0%; Pred. No. 3.2e-12;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNPNFASNDYSYWL 25
Db 1499 TMPFMFCNNVNCNPNFASNDYSYWL 1523

RESULT 4
CA14 MOUSE
ID CA14 MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1996 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83197932; PubMed=2703490;
RA Muthukuman G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagens.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.

```

```

RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbello P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM) forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; J03758; AAA37439.1; -
DR ENBL; M23333; AAA51625.1; -
DR ENBL; J04694; AAA50292.1; -
DR ENBL; X06777; CAA29946.1; -
DR ENBL; X02201; CAA36132.1; -
DR ENBL; M15832; AAA37340.1; -
DR ENBL; M14042; AAA37342.1; -
DR ENBL; M12879; AAA37343.1; -
DR ENBL; M13024; -; NOT_ANNOTATED_CDS.
DR ENBL; M13025; -; NOT_ANNOTATED_CDS.
DR ENBL; M13026; AAA37344.1; -
DR ENBL; M13027; AAA37345.1; -
DR ENBL; M13028; AAA37346.1; -
DR ENBL; J04448; AAA37437.1; -
DR PIR; A33525; CGMS4B.
DR MGD; MGI:188454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.

```

```
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Repeat; Hydroxylation; Connective tissue; Basement membrane;
KW Extracellular matrix; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 E -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E5205899 CRC64; .

Query Match 98.0%; Score 145; DB 1; Length 1669;
Best Local Similarity 96.0%; Pred. No. 3.2e-12;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPEMFCNNVNCNFSRNDYSYWL 25
DB 1499 TMPEMFCNNVNCNFSRNDYSYWL 1523

RESULT 5
CA34 BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1] _
RN TISSUE=Lens.
RP SEQUENCE FROM N.A.
RC TISSUE=Lens.
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.C., Redders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RT encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=23118822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;

RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire',
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC EMBL; M63139; AAA62708.1; -.
DR PIR; A39024; A39024.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; ProcollagnC4.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match 94.6%; Score 140; DB 1; Length 471;
Best Local Similarity 92.0%; Pred. No. 4.5e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPEMFCNNVNCNFSRNDYSYWL 25
DB 300 TMPEMFCNNVNCNFSRNDYSYWL 324
```

```
RT RT "Alternative splicing of the NCI domain of the human alpha 3 (IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [9]
RN RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9337506;
RA Nomoto R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3 (IV) and
RT alpha4 (IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [10]
RN RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [11]
RN RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Reenders S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome.";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
CC the C-terminal part of the NCI domain;
CC Name=1;
CC IsoId=Q01955-1; Sequence=Displayed;
CC Name=2; Synonyms=V;
CC IsoId=Q01955-2; Sequence=VSP_001170;
CC Name=3; Synonyms=L5;
CC IsoId=Q01955-3; Sequence=VSP_001171;
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical VS domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC -!- DISEASE: Antibodies against the NCI domain of alpha3(IV) mediate
CC the autoimmune disease Goodpasture syndrome (MIM:233450), which is
CC characterized by hematuria and pulmonary hemorrhage.
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive
CC Alport syndrome (AS) (MIM:203780), an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC -----  
 DR EMBL; X80031; CAA56335.1; --  
 DR EMBL; AJ288487; CAC36101.1; JOINED.  
 DR EMBL; AJ288488; CAC36101.1; JOINED.  
 DR EMBL; AJ288489; CAC36101.1; JOINED.  
 DR EMBL; AJ288490; CAC36101.1; JOINED.  
 DR EMBL; AJ288491; CAC36101.1; JOINED.  
 DR EMBL; AJ288492; CAC36101.1; JOINED.  
 DR EMBL; AJ288493; CAC36101.1; JOINED.  
 DR EMBL; AJ288494; CAC36101.1; JOINED.  
 DR EMBL; AJ288495; CAC36101.1; JOINED.  
 DR EMBL; AJ288496; CAC36101.1; JOINED.  
 DR EMBL; AJ288497; CAC36101.1; JOINED.  
 DR EMBL; AJ288498; CAC36101.1; JOINED.  
 DR EMBL; AJ288499; CAC36101.1; JOINED.  
 DR EMBL; AJ288500; CAC36101.1; JOINED.  
 DR EMBL; AJ288501; CAC36101.1; JOINED.  
 DR EMBL; AJ288502; CAC36101.1; JOINED.  
 DR EMBL; AJ288503; CAC36101.1; JOINED.  
 DR EMBL; AJ288504; CAC36101.1; JOINED.  
 DR EMBL; AJ288505; CAC36101.1; JOINED.  
 DR EMBL; AJ288506; CAC36101.1; JOINED.  
 DR EMBL; AJ288507; CAC36101.1; JOINED.  
 DR EMBL; AJ288508; CAC36101.1; JOINED.  
 DR EMBL; AJ288509; CAC36101.1; JOINED.  
 DR EMBL; AJ288510; CAC36101.1; JOINED.  
 DR EMBL; AJ288511; CAC36101.1; JOINED.  
 DR EMBL; AJ288512; CAC36101.1; JOINED.  
 DR EMBL; AJ288513; CAC36101.1; JOINED.  
 DR EMBL; AJ288514; CAC36101.1; JOINED.  
 DR EMBL; AJ288515; CAC36101.1; JOINED.  
 DR EMBL; AJ288516; CAC36101.1; JOINED.  
 DR EMBL; AJ288517; CAC36101.1; JOINED.  
 DR EMBL; AJ288518; CAC36101.1; JOINED.  
 DR EMBL; AJ288519; CAC36101.1; JOINED.  
 DR EMBL; AJ288520; CAC36101.1; JOINED.  
 DR EMBL; AJ288521; CAC36101.1; JOINED.  
 DR EMBL; AJ288522; CAC36101.1; JOINED.  
 DR EMBL; AJ288523; CAC36101.1; JOINED.  
 DR EMBL; AJ288524; CAC36101.1; JOINED.  
 DR EMBL; AJ288525; CAC36101.1; JOINED.  
 DR EMBL; AJ288526; CAC36101.1; JOINED.  
 DR EMBL; AJ288527; CAC36101.1; JOINED.  
 DR EMBL; AJ288528; CAC36101.1; JOINED.  
 DR EMBL; AJ288529; CAC36101.1; JOINED.  
 DR EMBL; AJ288530; CAC36101.1; JOINED.  
 DR EMBL; AJ288531; CAC36101.1; JOINED.  
 DR EMBL; AJ288532; CAC36101.1; JOINED.  
 DR EMBL; AJ288533; CAC36101.1; JOINED.  
 DR EMBL; AJ288534; CAC36101.1; JOINED.  
 DR EMBL; AJ288535; CAC36101.1; JOINED.  
 DR EMBL; AJ288536; CAC36101.1; JOINED.

Query Match 93.9%; Score 139; DB 1; Length 1670;  
 Best Local Similarity 88.0%; Pred. No. 2.2e-11;  
 Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFPCNNVNCVFASRNDYSYL 25  
 |||:||||:|||||  
 DB 1499 TWPFFPCNNVNCVFASRNDYSYL 1523

RESULT 7  
 CA24\_ACSU STANDARD; PRT; 1763 AA.  
 ID CA24\_ACSU  
 AC P27393;  
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, last sequence update)  
 DT 15-MAR-2004 (Rel. 43, last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 OS Ascaris suum (pig roundworm) (Ascaris lumbricoidea).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
 OC Ascarididae; Ascaris.  
 OX NCBI\_TaxID=6253;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).  
 RA MEDLINE=91340768; PubMed=1714907;  
 RA Pettitt J., Kingston I.B.;  
 RT "The complete primary structure of a nematode alpha 2(IV) collagen  
 RT and the partial structural organization of its gene";  
 RL J. Biol. Chem. 266:16149-16156(1991).  
 CC -!- FUNCTION: Collagen type IV is specific for basement membranes.  
 CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
 CC Type IV collagen forms a mesh-like network linked through  
 CC intermolecular interactions between 7S domains and between NC1  
 CC domains.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=I;  
 CC IsoId=P27393-1; Sequence=Displayed;  
 CC Name=II;  
 CC IsoId=P27393-2; Sequence=VSP 001159;  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the  
 CC G-X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL; M67507; AAA18014.1; -  
 DR PIR; S16366; S16366.  
 DR InterPro; IPR008161; Clg helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 25.  
 DR ProDom; PD000007; Clg helix; 6.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;  
 DR Alternative splicing; Glycoprotein; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 27 42 7S DOMAIN.  
 FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1530 1763 NONHELICAL REGION (NC1).  
 FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).  
 FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).  
 FT DISULFID 1593 1599 BY SIMILARITY.  
 FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).  
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).  
 FT DISULFID 1702 1709 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 249 249 O-LINKED (XYL. .) (GLYCOSAMINOGLYCAN  
 FT (IN ISOFORM II) (POTENTIAL)  
 FT VARSPLIC 230 266 GSGPFGPGPGPVPSTGAKTIIGPEGAPGNKGEK ->  
 FT GDIGPAGPGPGPFPSTGSGSIVGRHSGDKGVK (in

```
FT isoform II).
FT /FTID=VSP 001159.
SQ SEQUENCE 1763 AA; 168526 MW; 304F528BC06AAEOD CRC64;
Query Match 85.8%; Score 127; DB 1; Length 1763;
Best Local Similarity 80.0%; Pred. No. 1e-09; Indels 0; Gaps 0;
Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TMPPMFCNNVNCVFASRNDYSYWL 25
Db 1587 TMPPFLCDVNNVNCVFASRNDYSYWL 1611

RESULT 8
CA14_CABEL STANDARD; PRT; 1758 AA.
AC PL1139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CLB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RT collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnson L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lighthouse J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and through NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
```



```

DR EMBL; U22327; AAA64312.1; ALT_SEQ.
DR EMBL; U53342; AAA96215.1; -.
DR EMBL; U53342; AAA96216.1; -.
DR PIR; T29350; T29350.
DR WormPep; F01G12.5a; CB04334.
DR WormPep; F01G12.5b; CB04335.
DR GO; GO:0005587; C:collagen type IV; IMP.
DR GO; GO:0030020; P:extracellular matrix structural constituent. .; IMP.
DR GO; GO:0016043; P:cell organization and biogenesis; NAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; clg_helix; 6.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMO0111; C4; 2.
DR KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen; Alternative splicing; Glycoprotein; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1758 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 27 42 7S DOMAIN.
FT DOMAIN 42 1527 TRIPLE-HELICAL REGION.
FT DOMAIN 1528 1758 NONHELICAL REGION (NCL).
FT DISULFID 1546 1635 OR 1632 (BY SIMILARITY).
FT DISULFID 1579 1632 OR 1635 (BY SIMILARITY).
FT DISULFID 1591 1597 BY SIMILARITY.
FT DISULFID 1594 1750 OR 1747 (BY SIMILARITY).
FT DISULFID 1588 1747 BY SIMILARITY.
FT DISULFID 1700 1707 BY SIMILARITY.
FT CARBOHYD 248 248 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT VARSPLIC 229 264 GDIGSGVGPFGPPGPGPTGGSTGVSPGRNGPKGDK -> G
DIAGAPGAPGPPGPIASTWKSGTIIGFKGDLGEKGKEK (in isoform II).
/ETId=VSP_001160.
G->E (IN WN114; 73% LETHAL).
A->T (IN WN126; 100% LETHAL).
G->E (IN WN109; 37% LETHAL).
G->R (IN WN103 AND WN151; 96% LETHAL).
G->R (IN WN101; 100% LETHAL).
G->E (IN WN129; 100% LETHAL).
G->E (IN WN143; 100% LETHAL).
G->R (IN GN30; 90% LETHAL).
G->R (IN EL470; 94% LETHAL).
G->E (IN WN139; 20% LETHAL).
G->D (IN GN25; 2% LETHAL).
G->D (IN GN147; 7% LETHAL).
G->D (IN GN37 AND B246; 9% LETHAL).
E->D (IN REF. 3).
P->L (IN REF. 1 AND 3; AAA96216).
FW->P 167750 MW; 97EE3F3DBB2DA2CS CRC64;
Query Match 81.1%; Score 120; DB 1; Length 1758;
Best Local Similarity 76.0%; Pred.No. 9.5e-09;
Matches 19; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TMFMFCNNNVCMFASRNDSYWL 25
|||||:||||:||||:||||:
Db 1585 TMFFLCDFNNVCYNASNEKSYWL 1609
RESULT 10
CA64_HUMAN
AC CA64 HUMAN STANDARD; PRT; 1691 AA.
IC Q14031; Q12823; Q14053; Q9NQMS; Q9NTX3; Q9UW76; Q9DWG6; Q9Y4L4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RX MEDLINE=94171779; PubMed=8125972;  
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;  
RT Identification of a new collagen IV chain, alpha 6(IV), by cDNA  
RT isolation and assignment of the gene to chromosome Xq22, which is the  
RT same locus for COL4A5.";  
RL J. Biol. Chem. 269:7520-7526(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=94230418; PubMed=8175748;  
RA Zhou J., Ding M., Zhao Z., Redders S.T.;  
RT "Complete primary structure of the sixth chain of human basement  
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)  
RT and comparison with five other type IV collagen chains.";  
RL J. Biol. Chem. 269:13193-13199(1994).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND  
RP LYS-1110.  
RX MEDLINE=96299642; PubMed=8661006;  
RA Zhang X., Zhou J., Redders S.T., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Bird C., Grafham D., Lawlor S., Wilson S.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=93361972; PubMed=8356449;  
RA Zhou J., Kuchizuki T., Smeets H., Antignac C., Laurila P.,  
RA de Paeppe A., Tryggvason K., Redders S.T.;  
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RT inherited smooth muscle tumors.";  
RL Science 261:1167-1169(1993).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM) forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequence=VSP 001174;  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D21337; BRA04809.1; -  
CC EMBL; U04845; AAB19569.2; -  
CC EMBL; U47004; AAB19038.1; -  
CC EMBL; U46959; AAB19038.1; JOINED.  
CC EMBL; U46961; AAB19038.1; JOINED.  
CC EMBL; U46963; AAB19038.1; JOINED.  
CC EMBL; U46962; AAB19038.1; JOINED.  
CC EMBL; U46964; AAB19038.1; JOINED.  
CC EMBL; U46965; AAB19038.1; JOINED.  
CC EMBL; U46966; AAB19038.1; JOINED.  
CC EMBL; U46967; AAB19038.1; JOINED.  
CC EMBL; U46968; AAB19038.1; JOINED.  
CC EMBL; U46969; AAB19038.1; JOINED.  
CC EMBL; U46970; AAB19038.1; JOINED.  
CC EMBL; U46971; AAB19038.1; JOINED.  
CC EMBL; U46972; AAB19038.1; JOINED.  
CC EMBL; U46973; AAB19038.1; JOINED.  
CC EMBL; U46974; AAB19038.1; JOINED.  
CC EMBL; U46975; AAB19038.1; JOINED.  
CC EMBL; U46976; AAB19038.1; JOINED.  
CC EMBL; U46977; AAB19038.1; JOINED.  
CC EMBL; U46978; AAB19038.1; JOINED.  
CC EMBL; U46979; AAB19038.1; JOINED.  
CC EMBL; U46980; AAB19038.1; JOINED.  
CC EMBL; U46981; AAB19038.1; JOINED.  
CC EMBL; U46982; AAB19038.1; JOINED.  
CC EMBL; U46983; AAB19038.1; JOINED.  
CC EMBL; U46984; AAB19038.1; JOINED.  
CC EMBL; U46985; AAB19038.1; JOINED.  
CC EMBL; U46986; AAB19038.1; JOINED.  
CC EMBL; U46987; AAB19038.1; JOINED.  
CC EMBL; U46988; AAB19038.1; JOINED.  
CC EMBL; U46989; AAB19038.1; JOINED.  
CC EMBL; U46990; AAB19038.1; JOINED.  
CC EMBL; U46991; AAB19038.1; JOINED.  
CC EMBL; U46992; AAB19038.1; JOINED.  
CC EMBL; U46993; AAB19038.1; JOINED.  
CC EMBL; U46994; AAB19038.1; JOINED.  
CC EMBL; U46995; AAB19038.1; JOINED.  
CC EMBL; U46996; AAB19038.1; JOINED.  
CC EMBL; U46997; AAB19038.1; JOINED.  
CC EMBL; U46998; AAB19038.1; JOINED.  
CC EMBL; U46999; AAB19038.1; JOINED.  
CC EMBL; U47000; AAB19038.1; JOINED.  
CC EMBL; U47001; AAB19038.1; JOINED.  
CC EMBL; U47002; AAB19038.1; JOINED.  
CC EMBL; U47003; AAB19038.1; JOINED.  
CC EMBL; U47004; AAB19038.1; JOINED.  
CC EMBL; U46960; AAB19039.1; JOINED.  
CC EMBL; U46961; AAB19039.1; JOINED.  
CC EMBL; U46962; AAB19039.1; JOINED.  
CC EMBL; U46963; AAB19039.1; JOINED.  
CC EMBL; U46964; AAB19039.1; JOINED.  
CC EMBL; U46965; AAB19039.1; JOINED.  
CC EMBL; U46966; AAB19039.1; JOINED.  
CC EMBL; U46967; AAB19039.1; JOINED.  
CC EMBL; U46968; AAB19039.1; JOINED.  
CC EMBL; U46969; AAB19039.1; JOINED.  
CC EMBL; U46970; AAB19039.1; JOINED.  
CC EMBL; U46971; AAB19039.1; JOINED.  
CC EMBL; U46972; AAB19039.1; JOINED.  
CC EMBL; U46973; AAB19039.1; JOINED.  
CC EMBL; U46974; AAB19039.1; JOINED.  
CC EMBL; U46975; AAB19039.1; JOINED.  
CC EMBL; U46976; AAB19039.1; JOINED.  
CC EMBL; U46977; AAB19039.1; JOINED.  
CC EMBL; U46978; AAB19039.1; JOINED.  
CC EMBL; U46979; AAB19039.1; JOINED.  
CC EMBL; U46980; AAB19039.1; JOINED.  
CC EMBL; U46981; AAB19039.1; JOINED.  
CC EMBL; U46982; AAB19039.1; JOINED.  
CC EMBL; U46983; AAB19039.1; JOINED.

```

DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U46999; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -.
DR EMBL; AL105943; CAB96748.1; -.
DR EMBL; AL136080; CAB96748.1; -.
DR EMBL; AL031177; CAA20120.1; -.
DR EMBL; L22763; ABA16338.1; -.
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. ; NAS.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46
Query Match 76.4%; Score 113; DB 1; Length 1691;
Best Local Similarity 72.0%; Pred No. 8.5e-08;
Matches 18; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 TMPEFMCNINNVCFASRNDYSYL 25
||||:|||||:|||||:|||||
Db 1521 TMPEFMCNINNVCFASRNDYSYL 1545

RESULT 11
CA44 RABIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;

RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
alpha 4 chain of basement membrane collagen type IV and assignment of
the gene to the distal long arm of human chromosome 2.";
J. Biol. Chem. 267:23753-23758(1992).
-!- FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
alpha 6(IV), each of which can form a triple helix structure with
2 other chains to generate type IV collagen network.
-!- SUBCELLULAR LOCATION: Cell surface (potential).
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the G-
X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these are located in the NC1 domain, are conserved in all known type
IV collagens.
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L01477; -; NOT ANNOTATED_CDS.
DR PIR; A45137; A45137.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NONHELICAL REGION (NCL).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).
FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 464 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CBC9BB31242FE82 CRC64;

Query Match 70.3%; Score 104; DB 1; Length 623;
Best Local Similarity 64.0%; Pred No. 5.5e-07;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TMPEFMCNINNVCFASRNDYSYL 25
||||:|||||:|||||:|||||
Db 452 TLFPAYCNINHQVCHYAQRNDKSYWL 476

RESULT 12
CA44 HUMAN STANDARD; PRT; 1690 AA.
ID CA44 HUMAN
AC P53420;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 4(IV) chain precursor.
GN COL4A4.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=95014445; PubMed=7523402;  
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reiders S.T.;  
 RT "Complete primary structure of the human type IV collagen alpha 4 (IV)  
 RT chain. Comparison with structure and expression of the other alpha  
 RT (IV) chains.";  
 RL J. Biol. Chem. 269:26172-26177(1994).  
 RN [2]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=98198854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshiooka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [3]  
 RP SEQUENCE OF 1219-1690 FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=93374047; PubMed=8365481;  
 RA Sugimoto M., Ohashi T., Yoshiooka H., Matsuo N., Ninomiya Y.;  
 RT "cDNA isolation and partial gene structure of the human alpha 4 (IV)  
 RT collagen chain.";  
 RL FEBS Lett. 330:122-128(1993).  
 RN [4]  
 RP SEQUENCE OF 1407-1507 FROM N.A.  
 RX MEDLINE=93054733; PubMed=1429714;  
 RA Kanagata Y., Mattei M.-G., Ninomiya Y.;  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 RT alpha 4 chain of basement membrane collagen type IV and assignment of  
 RT the gene to the distal long arm of human chromosome 2.";  
 RL J. Biol. Chem. 267:23753-23758(1992).  
 RN [5]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [6]  
 RP VARIANT AS SER-1201.  
 RX MEDLINE=95078927; PubMed=7987396;  
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
 RA Smeets H.J.M., Reiders S.T.;  
 RT "Identification of mutations in the alpha 3 (IV) and alpha 4 (IV)  
 RT collagen genes in autosomal recessive Alport syndrome.";  
 RL Nat. Genet. 8:77-82(1994).  
 RN [7]  
 RP VARIANT FBH GLU-897.  
 RX MEDLINE=96379660; PubMed=8787673;  
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
 RT "Benign familial hematuria due to mutation of the type IV collagen  
 RT alpha4 gene.";  
 RL J. Clin. Invest. 98:1114-1118(1996).  
 RN [8]  
 RP VARIANTS AS, AND VARIANTS.  
 RX MEDLINE=99011253; PubMed=9792860;  
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,  
 RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;  
 RT "Determination of the genomic structure of the COL4A4 gene and of  
 RT novel mutations causing autosomal recessive Alport syndrome.";  
 RL Am. J. Hum. Genet. 63:1329-1340(1998).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -

CC alpha 6(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
 CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
 CC colocalized and present only in basement membranes of kidney, eye,  
 CC cochlea, lung and brain.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -!- DISEASE: Defects in COL4A4 are a cause of autosomal recessive  
 CC Alport syndrome (AS) [MIM:203780], an hereditary disorder  
 CC characterized by progressive glomerulonephritis, renal failure,  
 CC hematuria, ocular abnormalities and deafness. The recessive form  
 CC occurs equally between males and females.  
 CC -!- DISEASE: Defects in COL4A4 are a cause of familial benign  
 CC hematuria (FBH) [MIM:141200] or thin basement membrane disease.  
 CC FBH is characterized by persistent hematuria, an electron  
 CC microscopically detectable thin glomerular basement membrane (GBM)  
 CC and an autosomal dominant mode of inheritance. Renal function  
 CC remains normal. In children, differentiation between FBH and AS  
 CC can be difficult, because both disorders are manifested by  
 CC persistent hematuria and thin GBM at that age.  
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X81053; CAA56943.1; -.  
 CC EMBL; AB008496; BAA25065.1; -.  
 CC EMBL; D17391; BAA04214.1; -.  
 CC FIRM; A55360; CGHUIB.  
 CC GENE; HGNC:2206; COL4A4.  
 CC MIM; 120131; -.  
 CC MIM; 141200; -.  
 CC MIM; 203780; -.  
 CC InterPro; IPR008161; C1g helix.  
 CC InterPro; IPR008160; Collagen.  
 CC InterPro; IPR001442; Procollagen4\_C.  
 CC Pfam; PF01413; C4; 2.  
 CC Pfam; PF01391; Collagen; 21.  
 CC ProDom; PD000007; C1g helix; 3.  
 CC ProDom; PD003923; ProcollagenC4; 1.  
 CC SMART; SM00111; C4; 2.  
 CC Extracellular matrix; Connective tissue; Basement membrane; Repeat;  
 CC Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;  
 CC Polymorphism; Alport syndrome.  
 CC SIGNAL 1 38  
 CC CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.  
 CC DOMAIN 39 64 7S DOMAIN.  
 CC DOMAIN 65 1459 TRIPLE-HELICAL REGION.  
 CC SITE 94 96 NONHELICAL REGION (NC1).  
 CC SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).  
 CC SITE 1206 1207 CELLEAVAGE (BY COLLAGENASE)  
 CC (BY SIMILARITY).  
 CC FT



```

DR EMBL; X04410; CAA27998.1; -
DR EMBL; X02896; CAA26655.1; -
DR EMBL; X02897; CAA51614.1; -
DR EMBL; X02898; CAA26657.1; -
DR EMBL; X02899; CAA26658.1; -
DR PIR; A33526; A33526; -
DR MGI; MGI:88455; Col4a2; -
DR GO; GO:0005604; C:basement membrane; IDA; -
DR InterPro; IPR008161; Clg_helix; -
DR InterPro; IPR008160; Collagen; -
DR InterPro; IPR001442; Procollagn4_C; -
DR Pfam; PF01413; C4; 2; -
DR ProDom; PD000007; Clg_helix; 7; -
DR ProDom; PD000923; ProcollagnC4; 1; -
DR SMART; SM00111; C4; 2; -
DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT PROPEP 26 183 COLLAGEN ALPHA 2 (IV) CHAIN.
FT CHAIN 184 1707 TRIPLE-HELICAL REGION.
FT DOMAIN 1480 1479 NONHELICAL REGION (NC1).
FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).
FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).
FT DISULFID 1544 1550 BY SIMILARITY.
FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).
FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).
FT DISULFID 1653 1660 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1051 1051 P -> R (IN REF. 6).
FT CONFLICT 1097 1097 S -> G (IN REF. 7).
FT CONFLICT 1171 1171 G -> S (IN REF. 6).
FT CONFLICT 1179 1179 P -> R (IN REF. 6).
FT CONFLICT 1241 1241 Q -> E (IN REF. 6).
FT CONFLICT 1328 1328 P -> A (IN REF. 6).
FT CONFLICT 1573 1573 V -> L (IN REF. 4).
FT CONFLICT 1623 1623 Y -> H (IN REF. 4).
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query Match 70.3%; Score 104; DB 1; Length 1707;
Best Local Similarity 68.0%; Pred. No. 1.5e-06;
Matches 17; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 TMPEFMFCNNVNCNFSRNDYSYWL 25
Db 1538 TMPEFLCNPDCVCYASRNDKSYWL 1562
|||||:||||:|||||:|||||
|||||:||||:|||||:|||||

RESULT 14
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2 (IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19488-19493(1988).
RN [2]
RP SEQUENCE OF 1-1042 FROM N.A.
RX TISSUE=Placenta;

```

```

RX MEDLINE=88151998; PubMed=3345760;
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence
of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
chain reveals deletions in the alpha 1(IV) chain.";
RL Eur. J. Biochem. 172:35-42(1988).
RN [3]
RP SEQUENCE OF 1254-1712 FROM N.A.
RX MEDLINE=87219158; PubMed=3592677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
(IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286(1987).
RN [4]
RP SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
RN [5]
RP SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
collagen are divergently encoded on opposite DNA strands and have an
overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [7]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
basement membrane collagen type IV are arranged head-to-head and
separated by a bidirectional promoter of unique structure.";
RL EMBO J. 7:2687-2695(1988).
RN [8]
RP SEQUENCE OF 1-33 FROM N.A.
RX TISSUE=Skin;
RA MEDLINE=93305049; PubMed=8317999;
RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter
region of human collagen type IV genes, involved in the regulation of
divergent transcription.";
RL Biochem. J. 292:687-695(1993).
RN [9]
RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RX TISSUE=Placenta;
RA MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
carboxyterminal, non-collagenous aggregation and cross-linking domain
of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
RN [10]
RP FUNCTION: Type IV collagen is the major structural component of
glomerular basement membranes (GBM), forming a 'chicken-wire'
meshwork together with laminins, proteoglycans and entactin/
nidogen.
RN [11]
RP SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
alpha 6(IV), each of which can form a triple helix structure
with 2 other chains to generate type IV collagen network.
RN [12]
RP DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal

```





```
FT NON_TER 1 1
FT DOMAIN <1 222
FT DOMAIN 223 453
FT DISULFID 243 332
FT DISULFID 276 329
FT DISULFID 288 294
FT DISULFID 351 449
FT DISULFID 385 446
FT DISULFID 397 404
FT CONFLICT 219 219
SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

      TRIPLE-HELICAL REGION.
      NONHELICAL REGION (NC1).
      OR 329 (BY SIMILARITY).
      OR 332 (BY SIMILARITY).
      BY SIMILARITY.
      OR 446 (BY SIMILARITY).
      OR 449 (BY SIMILARITY).
      BY SIMILARITY.
      I -> P (IN REF. 2 AND 3).
      Query Match 69.6%; Score 103; DB 1; Length 453;
      Best Local Similarity 64.0%; Pred No. 5, 5e-07;
      Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 1 TMPFMFCNINNVCFASRNDYSYL 25
DB 282 TLPFAYCNIHQVCHYARRNDRSYWL 306
```

Search completed: April 5, 2004, 06:59:38  
Job time : 3.1477 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 15.0121 Seconds  
(without alignments)  
525.440 Million cell updates/sec

Title: US-10-032-221b-38  
Perfect score: 148  
Sequence: 1 TMPEMFCNNVNCNPNRNDYSYWL 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacterioph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	179	11 P70165	P70165 mus musculus
2	148	100.0	253	11 Q61436	Q61436 mus musculus
3	148	100.0	585	11 Q80V57	Q80V57 mus musculus
4	148	100.0	799	11 Q8NS7	Q8NS7 mus musculus
5	148	100.0	886	4 Q9NUE7	Q9NUE7 homo sapien
6	148	100.0	1684	6 Q8HYC1	Q8HYC1 canis fami
7	148	100.0	1688	6 Q86622	Q86622 canis fami
8	148	100.0	1691	11 Q9ESQ2	Q9ESQ2 mus musculus
9	145	98.0	161	11 Q61430	Q61430 mus musculus
10	145	98.0	210	6 Q28273	Q28273 canis fami
11	145	98.0	225	6 Q28271	Q28271 canis fami
12	145	98.0	228	11 Q9SLQ8	Q9SLQ8 mus musculus
13	145	98.0	229	4 Q8NF88	Q8NF88 homo sapien
14	145	98.0	229	4 Q9NYC5	Q9NYC5 homo sapien
15	145	98.0	246	11 Q61435	Q61435 mus musculus
16	145	98.0	979	13 Q919K3	Q919K3 gallus gall

17	145	98.0	1075	4 Q86X41	Q86X41 homo sapien
18	145	98.0	1621	4 Q9HAR9	Q9HAR9 homo sapien
19	145	98.0	1669	11 Q9QZS0	Q9QZS0 mus musculus
20	140	94.6	203	6 Q29032	Q29032 sus scrofa
21	140	94.6	203	6 Q28682	Q28682 cryetolagus
22	140	94.6	212	6 Q28567	Q28567 ovis aries
23	139	93.9	212	6 Q28512	Q28512 macaca mula
24	139	93.9	230	11 Q63122	Q63122 rattus norv
25	139	93.9	245	4 Q9NYC4	Q9NYC4 homo sapien
26	139	93.9	1752	5 Q07265	Q07265 strongyloce
27	129	87.2	1747	5 Q26640	Q26640 strongyloce
28	121	81.8	1802	5 Q17163	Q17163 brugia mala
29	113	76.4	205	6 Q28274	Q28274 canis fami
30	113	76.4	546	11 Q99K97	Q99K97 mus musculu
31	113	76.4	1600	4 Q9UEH6	Q9UEH6 homo sapien
32	113	76.4	1691	11 Q9ESQ1	Q9ESQ1 mus musculu
33	108	73.0	1723	5 Q9GQB1	Q9GQB1 hydra atten
34	104	70.3	202	6 Q28272	Q28272 canis fami
35	104	70.3	312	11 Q64457	Q64457 mus musculu
36	104	70.3	358	11 Q91VI3	Q91VI3 mus musculu
37	104	70.3	673	4 Q14052	Q14052 homo sapien
38	104	70.3	1682	11 Q9QZ89	Q9QZ89 mus musculu
39	103	69.6	208	6 Q29468	Q29468 canis fami
40	96	64.9	713	5 Q9GV24	Q9GV24 sarcophaga
41	96	64.9	1024	5 Q8T7S4	Q8T7S4 anopheles g
42	96	64.9	1761	5 Q18407	Q18407 drosophila
43	96	64.9	1940	5 Q9VMV5	Q9VMV5 drosophila
44	95	64.2	1779	5 Q9VMV4	Q9VMV4 drosophila
45	69	46.6	854	5 Q09238	Q09238 pseudocorti

## ALIGNMENTS

## RESULT 1

P70165 PRELIMINARY; PRT; 179 AA.  
AC P70165;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha5 chain (Fragment).  
GN COL4A5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RA Oberbaumer I.;  
RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and alpha5(IV) in differentiated teratocarcinoma cells."  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X82218; CAA57698.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
FT NON\_TER 1  
FT NON\_TER 179 179  
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 100.0%; Score 148; DB 11; Length 179;  
Best Local Similarity 100.0%; Pred. No. 2.3e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPEMFCNNVNCNPNRNDYSYWL 25

DB 36 TMPEMFCNNVNCNPNRNDYSYWL 60

```
RESULT 2
Q61436
ID Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RN SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RC MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sares J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: Sequence distribution, association with laminins, and
RT developmental switches.";
RL J. Cell Biol. 127:875-891 (1994).
DR EMBL; 235168; CAA84531.1; -.
DR PIR; I48304; I48304.
DR MGD; MGI:88456; Col14a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match 100.0%; Score 148; DB 11; Length 253;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPEFMCNINNVCFASRNDYSYWL 25
Db 83 TMPEFMCNINNVCFASRNDYSYWL 107

RESULT 3
Q80V57
ID Q80V57 PRELIMINARY; PRT; 585 AA.
AC Q80V57;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Col4a5 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
```

```
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maiz M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RC Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043317; AAH43317.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
SQ SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 100.0%; Score 148; DB 11; Length 585;
Best Local Similarity 100.0%; Pred. No. 7.7e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPEFMCNINNVCFASRNDYSYWL 25
Db 415 TMPEFMCNINNVCFASRNDYSYWL 439

RESULT 4
Q8NS7
ID Q8NS7 PRELIMINARY; PRT; 799 AA.
AC Q8NS7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGD; MGI:88456; Col14a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 100.0%; Score 148; DB 11; Length 799;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 629 TMPFMFCNNVNCNPFASNDYSYWL 653

RESULT 5
Q9NUB7 PRELIMINARY; PRT; 886 AA.
ID Q9NUB7
AC Q9NUB7
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER
SQ SEQUENCE 886 AA; 85479 MW; 806B9FCA9AA6569 CRC64;

Query Match 100.0%; Score 148; DB 4; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 716 TMPFMFCNNVNCNPFASNDYSYWL 740

RESULT 6
Q8HYC1 PRELIMINARY; PRT; 1684 AA.
ID Q8HYC1
AC Q8HYC1
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5 chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA Jacobs R., Thorner P.S.;
RL "Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
transfer to smooth muscle restores expression of the alpha5(IV)
collagen chain in a canine model of Alport syndrome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY078501; AAL83712.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 1521 TMPFMFCNNVNCNPFASNDYSYWL 1545

RESULT 7
Q866Z2 PRELIMINARY; PRT; 1688 AA.
ID Q866Z2
AC Q866Z2
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5.
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Cox M.L., Lees G.E., Kashtan C.E., Murphy K.E.;
RL "Genetic Cause of X-linked Alport Syndrome in a Family of Domestic
Dogs.";
RL Mamm. Genome 0:0-0(2003).
DR EMBL; AF470624; AAC33458.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
RP SEQUENCE FROM N.A.
RA Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA Jacobs R., Thorner P.S.;
RL "Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
transfer to smooth muscle restores expression of the alpha5(IV)
collagen chain in a canine model of Alport syndrome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY078501; AAL83712.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.

QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25
Db 1521 TMPFMFCNNVNCNPFASNDYSYWL 1545

RESULT 8
Q9ESQ2 PRELIMINARY; PRT; 1691 AA.
ID Q9ESQ2
AC Q9ESQ2
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type IV collagen alpha 5 chain.
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20536494; PubMed=10965041;
RX Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
RA Kishiro Y., Sado Y., Yoshioka H., Ninomiya Y.,
RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in
RT Epithelial Basement Membranes.";
RL J. Biochem. 128:427-434(2000).
DR EMBL; AB041350; BAB13673.1; -.
DR GO; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 100.0%; Score 148; DB 11; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNINNVCFASRNDYSYWL 25
DB 1521 TMPFMFCNINNVCFASRNDYSYWL 1545

RESULT 9
Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Oberbauer I.;
RT "Cloning of the NCI domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 161;
Best Local Similarity 96.0%; Pred. No. 5.9e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNINNVCFASRNDYSYWL 25
```

```
DB 8 TMPFLFCNINNVCFASRNDYSYWL 32

RESULT 10
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA923633D CRC64;

Query Match 98.0%; Score 145; DB 6; Length 210;
Best Local Similarity 96.0%; Pred. No. 7.7e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNINNVCFASRNDYSYWL 25
DB 51 TMPFLFCNINNVCFASRNDYSYWL 75

RESULT 11
Q28271 PRELIMINARY; PRT; 225 AA.
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-Sanoyed;
RA Thorne P.S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50933; AAC49583.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON_TER 1 225
FT NON_TER 225
SQ SEQUENCE 225 AA; 24585 MW; 2C204558950416E47 CRC64;

Query Match 98.0%; Score 145; DB 6; Length 225;
Best Local Similarity 96.0%; Pred. No. 9.2e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 65 TMPFLFCNNVNCVFASRNDYSYWL 89

RESULT 12
Q99LQ8 PRELIMINARY; PRT; 226 AA.
AC Q99LQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002269; RAH02269.1; -.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Hypothetical Protein.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 226 AA; 25042 MW; 4F7FD5371181C21 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 226;
Best Local Similarity 96.0%; Pred. No. 8.3e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 56 TMPFLFCNNVNCVFASRNDYSYWL 80

RESULT 13
Q9NF88 PRELIMINARY; PRT; 229 AA.
AC Q9NF88;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN SEQUENCE FROM N.A.
RP He A.B.;
RT "Cloning and Expression of Arresten in Escherichia coli and Pachia pastoris.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF536207; AAM97359.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 229 AA; 25391 MW; 09B21FDSAB517E9E CRC64;

Query Match 98.0%; Score 145; DB 4; Length 229;
Best Local Similarity 96.0%; Pred. No. 8.4e-14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25
DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 14
Q9NYC5 PRELIMINARY; PRT; 229 AA.
AC Q9NYC5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H., Takahashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K., Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D., Weichselbaum R.R., Sukhatme V.P., Kalluri R.;
RT "Anti-angiogenic cues from vascular basement membrane collagen.";
RL Cancer Res. 0:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Fu J., Bai X., Wang W., Ruan C.;
RT "Arresten, a collagen-derived inhibitor of angiogenesis.";
RL Chung Hua Heueh Yeh Heueh Tea Chih 22:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Peng X., Yin B., Yuan J., Qiang B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;
RT "Molecular cloning and sequencing of human arresten gene.";
RL Zhonghua Shi Yan Wai Ke Za Zhi 19:46-47(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Song Z.F., Zheng Q.C.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258349; AAF72630.1; -.
DR EMBL; AF363672; AAK53382.1; -.
DR EMBL; AF400431; AAK92480.1; -.
DR EMBL; AY285780; AAP49112.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
```

FT NON TER 1 1  
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;  
Query Match 98.0%; Score 145; DB 4; Length 229;  
Best Local Similarity 96.0%; Pred. No. 8.4e-14;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMPFMFCNINNVCFASRNDYSYWL 25  
DB 59 TMPFMFCNINNVCFASRNDYSYWL 83

RESULT 15

O61435 PRELIMINARY; PRT; 246 AA.  
AC Q61435  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Collagen IV alpha 3 chain (Fragment).  
CN COL4A3  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/c;  
RX MEDLINE=95050957; PubMed=7962065;  
RA Miner J.H.; Sanes J.R.;  
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal  
RT laminae: Sequence, distribution, association with laminins, and  
RT developmental switches."  
RL J. Cell Biol. 127:879-891(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/c;  
RA Miner J.H.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z35166; CAA84529.1; -.  
DR PIR; I48302; I48302.  
DR MGI; MGI:104688; Col4a3  
DR GO; GO:0005604; C:basement membrane; IDA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RNP\_RNP\_1; 1.  
FT NON-TER 1  
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 98.0%; Score 145; DB 11; Length 246;  
Best Local Similarity 96.0%; Pred. No. 9e-14;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TMPFMFCNINNVCFASRNDYSYWL 25  
DB 75 TMPFMFCNINNVCFASRNDYSYWL 99

Search completed: April 5, 2004, 07:03:57  
Job time: 15.0121 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 22.5182 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221b-38

Perfect score: 148

Sequence: 1 TWFFMFCNNVNCNPNRNDYSYWL 25.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	25	6	ADA20237 T7 mutant
2	148	100.0	229	7	ADC17699 Human typ
3	148	100.0	284	2	AAY31995 Type IV c
4	148	100.0	264	3	AAY37557 Human alp
5	148	100.0	309	3	AAB54044 Human pan
6	148	100.0	772	2	AAR23873 Human alp
7	148	100.0	772	2	AAR23873 Human alp
8	148	100.0	1685	4	AAW09643 Human typ
9	148	100.0	1693	4	ABG04839 Novel hum
10	145	98.0	229	1	ABG15619 Novel hum
11	145	98.0	229	3	AAP3524 Complete
12	145	98.0	229	5	AAY67943 Human typ
13	145	98.0	229	6	AAY75587 Human typ
14	145	98.0	229	6	ADA20217 Human typ
15	145	98.0	229	6	ADC17695 Human typ
16	145	98.0	260	2	AAY31991 Type IV c
17	145	98.0	260	3	AAY37553 Human alp
18	145	98.0	406	3	AAB58169 Lung can
19	145	98.0	1669	4	AAM40863 Human pol
20	145	98.0	1669	5	ABB90760 Human Tum
21	145	98.0	1669	5	ABBS7334 Mouse isc
22	145	98.0	1669	6	ABUS4467 Human tum
23	140	94.6	471	2	AAM39077 Human pol
24	140	94.6	471	2	AAR79163 Partial s
25	140	94.6	471	3	AAY44171 Bovine ty
					AAY56783 Bovine al

26	140	94.6	471	4	AAE09483	Aae09483 Bovine al
27	139	93.9	25	6	ADA20236	Ada20236 T7 peptid
28	139	93.9	79	5	AAU75600	Aau75600 Human typ
29	139	93.9	79	5	ADA20264	Ada20264 Human tum
30	139	93.9	88	5	AAU75608	Aau75608 Human typ
31	139	93.9	88	5	AAU75607	Aau75607 Human typ
32	139	93.9	88	6	ADA20271	Ada20271 Human tum
33	139	93.9	88	6	ADA20272	Ada20272 Human tum
34	139	93.9	124	5	AAU75594	Aau75594 Human typ
35	139	93.9	124	6	ADA20258	Ada20258 Human tum
36	139	93.9	132	5	AAU75597	Aau75597 Human typ
37	139	93.9	132	6	ADA20261	Ada20261 Human tum
38	139	93.9	191	5	AAU75596	Aau75596 Human typ
39	139	93.9	191	6	ADA20260	Ada20260 Human tum
40	139	93.9	211	3	AAU755918	Aay95918 Human Goo
41	139	93.9	211	5	ABG79208	Abg79208 Human GP
42	139	93.9	218	2	AAR79164	Aar79164 Partial s
43	139	93.9	218	2	AAU44172	Aay44172 Human typ
44	139	93.9	218	3	AAU56784	Aay56784 Human alp
45	139	93.9	218	4	AAE09484	Aae09484 Human alp

## ALIGNMENTS

### RESULT 1

ADA20237

ID ADA20237 standard; peptide; 25 AA.

XX AC ADA20237;

XX DT 20-NOV-2003 (first entry)

XX DE T7 mutant peptide related to human type IV collagen and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NCI; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytosolic; gene therapy; T7 mutant peptide; mutant; mutatin; human; type IV collagen alpha 3 chain; tumstatin; human.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 5 /note= "Wild-type Leu substituted by Met"

FT Misc-difference 9 /note= "Wild-type Val substituted by Ile"

FT Misc-difference 11 /note= "Wild-type Asp substituted by Asn"

WO2003059257-A2.

PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX PT New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 61; Page 45; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the mutated T7 peptide of the  
 CC invention. The wild-type T7 peptide sequence is given in Seq ID37 (see  
 CC ADA20236) and this was derived from the amino acid sequence of tumatatin,  
 CC which in turn was derived from the amino acid sequence of human type IV  
 CC collagen alpha 3 chain.

XX Sequence 25 AA;

Query Match 100.0%; Score 148; DB 6; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25  
 Db 1 TMPFMFCNNVNCNFSRNDYSYWL 25

RESULT 2

ADCL7699

ID ADCL7699 standard; protein; 229 AA.

XX ADCL7699;

DT 18-DEC-2003 (first entry)

DE Human type IV collagen alpha 5 chain protein SEQ ID NO:306.

XX crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.

XX Homo sapiens.

XX W02003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

XX 29-OCT-2001; 2001US-0351289P.

XX 22-MAR-2002; 2002US-0366854P.

XX 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX (SUND/) SUNDARAMOORTHY M.

XX (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.

PS Disclosure; SEQ ID NO 306; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
 CC antipsoriatic, anti-anaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents an amino acid sequence which is used in the exemplification of  
 CC the present invention.

XX Sequence 229 AA;

Query Match 100.0%; Score 148; DB 7; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFSRNDYSYWL 25  
 Db 59 TMPFMFCNNVNCNFSRNDYSYWL 83

RESULT 3

AAV31995

ID AAV31995 standard; protein; 264 AA.

XX AAV31995;

XX 05-JAN-2000 (first entry)

XX Type IV collagen NC1 domain alpha-5 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;  
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
 KW rheumatoid arthritis; retinal neovascularization;  
 KW choroidal neovascularization; macular degeneration;  
 KW corneal neovascularization; retinopathy of prematurity;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidemic keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; soggren's; acne rosacea; phlyctenulosis;  
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;  
 KW ulcer; herpes simplex infection; Herpes zoster infection;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW systemic lupus; polyarteritis; Wegener's sarcooidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;  
 KW sarcooid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Bales disease; Bechets disease; myopia;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;



CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,  
CC sickle cell anaemia, sarcoïd, carotid obstructive disease, post-laser  
CC complications, chronic inflammation or psoriasis  
XX  
SQ Sequence 264 AA;

Query Match 100.0%; Score 148; DB 3; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.9e-13; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
|||||  
Db 94 TMEFMCNINNVCFASRNDYSYWL 118  
|||||

## RESULT 5

AAB54044  
ID AAB54044 standard; protein; 309 AA.

XX  
AC AAB54044;

XX  
DT 09-MAR-2001 (first entry)

XX  
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:496.

XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;  
KW diagnosis; identification; cytostatic; neuroprotective; nootropic;  
KW immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiast; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic; neural;  
KW immune system; muscular; reproductive; gastrointestinal; pulmonary;  
KW cardiovascular; renal; proliferative.

XX  
OS Homo sapiens.

XX  
PN W0200055320-A1.

XX  
PD 21-SEP-2000.

XX  
PF 08-MAR-2000; 2000WO-US005989.

XX  
PR 12-MAR-1999; 99US-0124270P.

XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI Rosen CA, Ruben SM;

XX  
DR WPI; 2000-579444/54.

XX  
DR N-PSDB; AAC98809.

XX  
FT New nucleic acid that is a pancreatic cancer antigen for preventing,  
FT treating, or ameliorating a medical condition, particular pancreatic  
FT cancer, or for use in assays for diagnosing a pathological condition.

XX  
PS Claim 11; Page 934-935; 1379pp; English.

XX  
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiast and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.

XX  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including

CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention  
XX  
SQ Sequence 309 AA;

Query Match 100.0%; Score 148; DB 3; Length 309;  
Best Local Similarity 100.0%; Pred. No. 2.2e-13; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
|||||  
Db 139 TMEFMCNINNVCFASRNDYSYWL 163  
|||||

## RESULT 6

AAR23873

ID AAR23873 standard; protein; 772 AA.

XX  
AC AAR23873;

XX  
DT 25-NOV-1992 (first entry)

XX  
DE Human alpha 5 (IV) of type IV collagen.

XX  
DE Mutations; Alport's syndrome; basement membranes; diabetes mellitus.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT Misc-difference 43..47 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 159..160 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 275..277 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 334..336 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 456..458 /note= "interruption in Gly-X-Y sequence"

XX  
US5114840-A.

XX  
PD 19-MAY-1992.

XX  
PF 07-JUL-1989; 89US-00377238.

XX  
PR 07-JUL-1989; 89US-00377238.

XX  
PA (TRYG/) TRYGGVASON K.

XX  
PI Tryggvason K, Hostikka SL;

XX  
DR WPI; 1992-192174/23.

XX  
DR N-PSDB; AAQ24551.

XX  
FT Isolation of DNA encoding alpha-5(IV) polypeptide of type IV collagen - to  
FT detect mutations in genes for alpha-5(IV) chain which produce genetic or  
FT acquired basement membrane disorders e.g. Alport's syndrome.  
XX  
PS Disclosure; Fig 2; 14pp; English.

XX  
CC The sequence is that of the alpha 5(IV) polypeptide chain of human type  
CC IV collagen, the major component of basement membranes. The protein  
CC contains the Gly-X-Y repeat coding sequence typical for collagenous  
CC proteins at one end and a typical NC-domain coding sequence at the other  
CC end. The sequence can be used to detect mutations in individual genes  
CC specific for this chain which can, directly or indirectly, produce  
CC several human diseases. It can also be used to determine genetic, e.g.  
CC Alport's syndrome, or acquired e.g. diabetes mellitus, disorders of the  
CC basement membrane, and as probes or antibodies against these nucleotide

CC sequences. Gene fragments generated through amplifications from human  
 CC genomic or cloned DNA can also be used for detection and analysis of  
 CC genes

XX SQ Sequence 772 AA;  
 Query Match 100.0%; Score 148; DB 2; Length 772;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25  
 DB 602 TMPFMFCNNVNCNPFASNDYSYWL 626

RESULT 7  
 AAW09643  
 ID AAW09643 standard; protein; 772 AA.

XX AC AAW09643;  
 DT 25-MAR-2003 (revised)  
 DT 16-JUN-1997 (first entry)  
 XX DE Human type IV collagen alpha-5.  
 XX KW Collagen alpha5(IV); basement membrane; Alport's syndrome; nephritis;  
 KW kidney; renal failure; antibody; diagnosis; COL4A5 gene; X chromosome.  
 XX OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Domain 1..543  
 FT /label= Collagenous domain  
 FT /note= "collagenous domain contains Gly-X-Y tripeptide  
 repeats, interrupted at positions 43-47, 159-160, 275-  
 276, 334-335, 456-459"  
 FT 544..772  
 FT /label= Non-collagenous\_domain  
 FT 742..751  
 FT /label= Immunogenic peptide  
 FT /note= "peptide used to raise diagnostic antibodies  
 (Claim 1)"

XX US5593900-A.  
 XX 14-JAN-1997.  
 XX 11-OCT-1994; 94US-00321084.  
 XX 07-JUL-1989; 89US-00377238.  
 XX 20-DEC-1990; 90US-00630563.  
 XX (TRYG/) TRYGGVASON K.  
 PA (HOST/) HOSTIKKA S L.  
 PA (HOYH/) HOYHTYA M.  
 XX Hostikka SL, Tryggvason K, Hoyhtya M;  
 WPI; 1997-099481/09.  
 DR N-PSDB; AAT47812.  
 XX New antibodies specific for human type IV collagen alpha5 chain - used to  
 PT detect absence of this chain in patients with renal failure.  
 XX Disclosure; Fig 2A-2B; 12pp; English.

XX The amino acid sequence of a portion (AAW09643) of the previously unknown  
 CC human type IV collagen chain, alpha5(IV), was deduced from cDNA clones  
 CC (see also AAT47812) ctd. using probes based on conserved sequences of  
 CC human alpha1(IV) and alpha2(IV) collagen chains and of the Drosophila  
 CC alpha(IV) chain. It includes a complete non-collagenous domain that shows  
 CC 83% identity with that of alpha1(IV) and 63% with that of the alpha2(IV)

CC chain. Mutations in the alpha5(IV) gene (COL4A5) are associated with  
 CC Alport's syndrome. Antibodies raised against a peptide (see also  
 CC AAW09644) specific to alpha5(IV) can be used in the diagnosis of basement  
 CC membrane disorders such as Alport's syndrome. (Updated on 25-MAR-2003 to  
 XX correct pF field.)

XX SQ Sequence 772 AA;  
 Query Match 100.0%; Score 148; DB 2; Length 772;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TMPFMFCNNVNCNPFASNDYSYWL 25  
 DB 602 TMPFMFCNNVNCNPFASNDYSYWL 626

RESULT 8  
 ABG04839  
 ID ABG04839 standard; protein; 1685 AA.

XX AC ABG04839;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #4830.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.

XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 XX 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 WPI; 2001-639362/73.  
 DR N-PSDB; AAS69026.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 20; SEQ ID NO 35198; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1685 AA;  
Query Match 100.0%; Score 148; DB 4; Length 1685;  
Best Local Similarity 100.0%; Pred. No. 1.4e-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TMPFMFCNNVNCNPFASRNDYSYWL 25  
DB 1515 TMPFMFCNNVNCNPFASRNDYSYWL 1539  
RESULT 9  
ABG15619  
ID ABG15619 standard; protein; 1693 AA.  
XX  
AC ABG15619;  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15610.  
XX  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
FN 11-OCT-2001.  
XX  
PD 30-MAR-2001; 2001WO-US008631.  
XX  
PF 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS79806.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 45978; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have application of mutations  
CC in diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic  
CC amino acid sequences of the invention. NOTE: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1693 AA;  
Query Match 100.0%; Score 148; DB 4; Length 1693;  
Best Local Similarity 100.0%; Pred. No. 1.4e-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TMPFMFCNNVNCNPFASRNDYSYWL 25  
DB 1523 TMPFMFCNNVNCNPFASRNDYSYWL 1547  
RESULT 10  
AAP93524  
ID AAP93524 standard; protein; 229 AA.  
XX  
AC AAP93524;  
XX  
DT 25-MAR-2003 (revised)  
DT 03-OCT-2002 (revised)  
DT 04-JUN-1990 (first entry)  
XX  
DE Complete sequence of the alpha-1-NCl domain of type IV collagen.  
XX  
DE Alpha-1-NCl domain; type IV collagen; cell adhesion; heparin;  
KW aortic endothelial cells; metastatic carcinoma M4 cells; rat fibroblasts;  
KW MM fibrosarcoma cells; C6 glioma cell; A431 breast carcinoma cells;  
KW wound healing; implant acceptance.  
XX  
OS Homo sapiens.  
XX  
FN WO8903392-A.  
XX  
PD 20-APR-1989.  
XX  
PF 20-AUG-1988; 88WO-US003023.  
PR 08-OCT-1987; 87US-00106858.  
XX  
XX (MINU ) MINNESOTA UNIVERSITY.  
XX  
XX Tsilbary EC;  
XX  
XX WPI; 1989-130015/17.  
XX  
XX Polypeptide(s) with type IV collagen activity - used to promote wound  
PT healing, implant acceptance and cellular attachment and inhibit malignant  
PT cells.  
XX  
XX Fig 2; page 1/12; 40pp; English.  
PS  
XX The peptides in the features table are claimed (Claim 1, p. 22). They  
CC were synthesised using the Merrifield solid phase method. Binding assays  
CC were carried out using peptides TS-1, TS-2 and TS-3. TS-1 promotes  
CC adhesions of aortic endothelial cells, metastatic carcinoma M4 cells,  
CC normal rat fibroblasts, MM fibrosarcoma cells, C6 glioma cells and A431  
CC breast carcinoma cells. TS-2 binds to type IV collagen, to heparin and  
CC promotes adhesion of the above cells. Peptides TS-1, TS-2 and TS-3 may be  
CC used to promote wound healing and implant acceptance, promote cellular  
CC attachment to culture substrata or inhibit the metastasis of malignant  
CC cells. They may be used to coat a prosthetic device. (Updated on 03-OCT-  
CC 2002 to add missing OS field.) (Updated on 25-MAR-2003 to correct PF  
CC field.) (Updated on 25-MAR-2003 to correct PI field.)  
XX

SQL Sequence 229 AA;

Query Match 98.0%; Score 145; DB 1; Length 229;  
Best Local Similarity 96.0%; Pred. No. 4.4e-13;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASNDYSYL 25  
DB 59 TMAPFCNNVNCNFCASNDYSYL 83

RESULT 11

AAV67943  
ID AAV67943 standard; protein; 229 AA.

XX  
AC AAV67943;

XX DT 03-APR-2000 (first entry)

XX DE Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.

XX KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;  
benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;  
ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;  
myocardial angiogenesis; plaque neovascularisation; angiofibroma;  
atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;  
contraception; obesity.

XX OS Homo sapiens.

XX PN WO9965940-A1.

XX PD 23-DEC-1999.

XX PF 17-JUN-1999; 99WO-US013737.

XX PR 17-JUN-1998; 98US-0089689P.

XX PR 25-MAR-1999; 99US-0126175P.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2000-097708/08.

XX DR N-PSDB; AAZ571159.

XX PT Anti-angiogenic proteins comprising the NCI domain of the alpha 1, 2 or 3  
chain of Type IV collagen used in, e.g. treatment of benign tumors and  
rheumatoid arthritis.

XX PS Example 1; Fig 1B; 117pp; English.

XX CC The present sequence represents the human type IV collagen alpha 1 chain.  
The present invention describes an isolated protein chosen from the NCI  
domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a  
fragment, analogue, derivative or mutant, which has anti-angiogenic  
properties. The anti-angiogenic proteins, multimers and chimeras are  
useful for inhibiting angiogenic activity in mammalian tissue, especially  
for treating diseases chosen from angiogenesis-dependent cancers, benign  
tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular  
angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,  
plaque neovascularisation, telangiectasia, haemophilic joints,  
angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori  
ulcers, dialysis graft vascular access stenosis, contraception and  
obesity. The compositions can be used to inhibit a disease characterised  
by angiogenic activity, in conjunction with radiation therapy,  
chemotherapy or immunotherapy

SQL Sequence 229 AA;

Query Match 98.0%; Score 145; DB 3; Length 229;  
Best Local Similarity 96.0%; Pred. No. 4.4e-13;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMAPFCNNVNCNFCASNDYSYL 25  
DB 59 TMAPFCNNVNCNFCASNDYSYL 83

RESULT 12

AAU75587  
ID AAU75587 standard; protein; 229 AA.

XX AC AAU75587;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 1 chain.

XX KW Human; type IV collagen alpha 1 chain; cytostatic; antiangiogenic;  
non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
Tumstatin; angiogenesis; tumour.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX DR N-PSDB; ABK15359.

XX PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and  
treating disorders involving angiogenesis.

XX PS Example 1; Fig 1B; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI  
domain, having one or more of the characteristics selected from: (a) the  
ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
proliferation of endothelial cells; and (c) the ability to cause  
apoptosis of endothelial cells. Also described are the following: (1) use  
of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
analogue or allelic variant in the preparation of a medicament for  
treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
where the angiogenesis is mediated by one or more endothelial cell  
integrins or one or more endothelial cell integrin subunits; or (b) by  
promoting or inducing endothelial cell apoptosis in a tissue, where the  
endothelial cell apoptosis is mediated by one or more endothelial cell  
integrins or one or more endothelial cell integrin subunits; (2) use of  
an antibody or peptide that specifically binds the alpha1, alpha2,  
alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the  
preparation of a medicament for inhibiting angiogenesis or cell  
proliferation; (3) use of an inhibitor, such as an antibody, antibody  
fragment or peptide of receptor-mediated angiogenesis in the preparation  
of a medicament for treating a proliferative disease in a vertebrate,  
where the disease is characterised by angiogenesis that is mediated by  
receptors to Arresten, Canstatin or Tumstatin and where the receptors  
inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
the presence of a medicament for promoting angiogenesis in a tissue; and  
(5) use of integrins in the preparation of a medicament for promoting or  
inducing angiogenesis or cell proliferation in a tissue. The fragments  
Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues



CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 1 chain  
 XX  
 SQ Sequence 229 AA;

Query Match 98.0%; Score 145; DB 5; Length 229;

Best Local Similarity 96.0%; Pred. No. 4.4e-13; Indels 0; Gaps 0;  
 Matches 24; Conservative 1; Mismatches 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25  
 Db 59 TMPFLFCNNVNCVFASRNDYSYWL 83

## RESULT 13

ADA20217  
 ID ADA20217 standard; protein; 229 AA.

XX AC ADA20217;  
 XX AC

DT 20-NOV-2003 (first entry)

XX Human type IV collagen alpha 1 chain partial protein sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytostatic; gene therapy; alpha 1 chain; arresten; human.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.

XX DR N-PSDB; ADA20216.

XX New peptide, useful for preparing a composition for inhibiting tumor  
 XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 101; Fig 1; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the partial amino acid sequence of the alpha 1 chain of human  
 CC type IV collagen. The "arresten" peptide of the invention was derived  
 CC from this protein.

XX Sequence 229 AA;

Query Match 98.0%; Score 145; DB 6; Length 229;

Best Local Similarity 96.0%; Pred. No. 4.4e-13;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25

Db 59 TMPFLFCNNVNCVFASRNDYSYWL 83

## RESULT 14

ADCI7695  
 ID ADCI7695 standard; protein; 229 AA.

XX AC ADCI7695;  
 XX AC

DT 18-DEC-2003 (first entry)

XX Human type IV collagen alpha 1 chain protein SEQ ID NO:302.

XX crystallised NC1 domain hexamer of type IV collagen;  
 XX angiogenesis inhibitor; angiogenesis-mediated disease;  
 XX tumour metastasis inhibitor; tumour growth inhibitor;  
 XX endothelial cell interaction inhibitor;  
 XX basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 XX anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 XX endothelial cell adhesion inhibitor;  
 XX endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 XX ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 XX blood-borne tumour.

XX OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX PA (SUND/) SUNDARAMOORTHY M.

XX PA (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 302; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina

CC membrane formation in cell or tissue development; (7) a crystal of an NCI  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and  
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents an amino acid sequence which is used in the exemplification of  
 CC the present invention.  
 XX  
 XX Sequence 229 AA;

Query Match 98.0%; Score 145; DB 7; Length 229;  
 Best Local Similarity 96.0%; Pred. No. 4.4e-13;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25  
 DB 59 TMPFLFCNNVNCVFASRNDYSYWL 83

RESULT 15  
 AAY31991  
 ID AAY31991 standard; protein; 260 AA.

XX AAY31991;  
 AC  
 XX  
 DT 05-JAN-2000 (first entry)  
 DE  
 DE Type IV collagen NCI domain alpha-1 monomer.

XX Type IV collagen; NCI domain; non-collagenous domain; human;  
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
 KW rheumatoid arthritis; retinal neovascularization;  
 KW choroidal neovascularization; retinopathy of prematurity;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidemic keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; soggrens; acne rosacea; phlyctenulosis;  
 KW syphilis; mycobacteria infection; lipid degeneration; chemical burn;  
 KW ulcer; herpes simplex infection; Herpes zoster infection;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW systemic lupus; polyarteritis; Wegener's sarcooidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;  
 KW sarcooid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Eales disease; Behcet's disease;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;  
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;  
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;  
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;  
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;  
 KW pemphigoid.

XX  
 OS Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers  
 FT Peptide 1..17  
 FT Protein /note= "BM40 signal peptide"  
 FT 18..260  
 FT /note= "mature protein"

FT Peptide 18..25  
 FT Protein /note= "affinity tag"  
 FT 26..260  
 FT /note= "NCI alpha-1 monomer"

PN WO9949885-A2.  
 XX 07-OCT-1999.  
 XX  
 XX 26-MAR-1999; 99WO-US006445.  
 XX  
 XX 27-MAR-1998; 98US-0079783P.  
 XX 29-OCT-1998; 98US-0106170P.  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Sarraz MP;  
 XX  
 XX WPI, 1999-601297/51.  
 DR N-PSDB; AAZ20089.

PT Inhibition of angiogenesis with non-collagenous alpha chain monomer  
 PT useful for treating e.g. tumor growth or metastasis, neovascularisation,  
 PT etc.

XX Disclosure; Fig 17a; 56pp; English.

XX This sequence represents a recombinant type IV collagen non-collagenous  
 CC (NCI) domain alpha-1 polypeptide composed of a BM40 signal sequence  
 CC (which is cleaved from the mature protein) to facilitate protein  
 CC secretion, and a mature protein comprising an affinity tag (facilitates  
 CC purification and identification of the material) and the alpha-1 chain  
 CC monomer. The invention provides methods and kits for inhibiting  
 CC angiogenesis, tumour growth and metastasis, and endothelial cell  
 CC interaction with the extracellular matrix, each method comprising  
 CC contacting the tumour or animal tissue with 1 or more isolated type IV  
 CC collagen NCI alpha chain monomer(s) selected from the group consisting of  
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NCI chain monomers (see AAY31991-  
 CC 96). The monomers can be produced via recombinant protein expression. The  
 CC polynucleotides and polypeptides are used to treat an angiogenesis-  
 CC mediated disorder or condition, especially selected from solid and blood-  
 CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal  
 CC neovascularization, choroidal neovascularization, macular degeneration,  
 CC corneal neovascularization, retinopathy of prematurity, corneal graft  
 CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic  
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic  
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, soggrens,  
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid  
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes  
 CC simplex infections, herpes zoster infections, protozoan infections,  
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal  
 CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's  
 CC sarcooidosis, scleritis, Steven's Johnson disease, radial keratotomy,  
 CC sickle cell anaemia, sarcooid, pseudoxanthoma elasticum, Pagets disease,  
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic  
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Behcet's  
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic  
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser  
 CC complications, abnormal proliferation of fibrovascular tissue,  
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,  
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative  
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)

XX Sequence 260 AA;

Query Match 98.0%; Score 145; DB 2; Length 260;  
 Best Local Similarity 96.0%; Pred. No. 5e-13;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCVFASRNDYSYWL 25  
 DB 90 TMPFLFCNNVNCVFASRNDYSYWL 114

Search completed: April 5, 2004, 06:58:31  
Job time : 22.5182 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 15.678 Seconds

(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-38

Perfect score: 148

Sequence: 1 TMSPFMCINNVCNFAFNDYSYWL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	100.0	229	14 US-10-032-221B-38	Sequence 38, Appl
2	148	100.0	229	14 US-10-032-221B-38	Sequence 306, App
3	148	100.0	309	9 US-09-925-297-496	Sequence 496, App
4	145	98.0	229	14 US-10-026-699-302	Sequence 302, App
5	145	98.0	229	14 US-10-032-221B-2	Sequence 2, Appl
6	145	98.0	406	9 US-09-925-302-507	Sequence 507, App
7	145	98.0	1669	15 US-10-372-683-8	Sequence 8, Appl
8	139	93.9	25	14 US-10-032-221B-37	Sequence 37, Appl
9	139	93.9	79	14 US-10-032-221B-26	Sequence 26, Appl
10	139	93.9	88	14 US-10-032-221B-33	Sequence 33, Appl
11	139	93.9	88	14 US-10-032-221B-33	Sequence 33, Appl
12	139	93.9	124	14 US-10-032-221B-20	Sequence 20, Appl
13	139	93.9	132	14 US-10-032-221B-23	Sequence 23, Appl
14	139	93.9	191	14 US-10-032-221B-22	Sequence 22, Appl
15	139	93.9	211	14 US-10-270-877-46	Sequence 46, Appl

16 139 93.9 211 14 US-10-270-837-46 Sequence 46, Appl  
17 139 93.9 232 14 US-10-206-699-304 Sequence 304, App  
18 139 93.9 244 14 US-10-032-221B-10 Sequence 10, Appl  
19 138 93.2 46 9 US-09-864-761-48095 Sequence 48095, A  
20 124 83.8 1759 15 US-10-369-493-7032 Sequence 7032, Ap  
21 121 81.8 1744 15 US-10-369-493-5832 Sequence 5832, Ap  
22 117 79.1 27 14 US-10-032-221B-39 Sequence 39, Appl  
23 115 77.7 22 14 US-10-206-699-267 Sequence 267, App  
24 113 76.4 142 9 US-09-864-761-38021 Sequence 38021, A  
25 113 76.4 228 14 US-10-206-699-307 Sequence 307, App  
26 112 75.7 22 14 US-10-206-699-265 Sequence 265, App  
27 106 71.6 22 14 US-10-206-699-266 Sequence 266, App  
28 105 70.9 18 14 US-10-206-699-261 Sequence 261, App  
29 104 70.3 227 14 US-10-206-699-303 Sequence 303, App  
30 104 70.3 227 14 US-10-032-221B-6 Sequence 6, Appli  
31 104 70.3 231 14 US-10-206-699-305 Sequence 305, App  
32 104 70.3 430 9 US-09-925-302-518 Sequence 518, App  
33 104 70.3 459 15 US-10-331-496A-27 Sequence 27, Appl  
34 104 70.3 459 15 US-10-372-683-30 Sequence 30, Appl  
35 104 70.3 1712 10 US-09-961-403-9 Sequence 9, Appli  
36 102 68.9 18 14 US-10-206-699-259 Sequence 259, App  
37 97 65.5 27 14 US-10-032-221B-40 Sequence 40, Appl  
38 96 64.9 18 14 US-10-206-699-260 Sequence 260, App  
39 94 63.5 18 14 US-10-206-699-255 Sequence 255, App  
40 93 62.8 27 14 US-10-032-221B-42 Sequence 42, Appl  
41 91 61.5 18 14 US-10-206-699-253 Sequence 253, App  
42 90 60.8 19 14 US-10-032-221B-41 Sequence 41, Appl  
43 90 60.8 20 14 US-10-206-699-290 Sequence 290, App  
44 89 60.1 15 14 US-10-206-699-210 Sequence 210, App  
45 89 60.1 22 14 US-10-206-699-270 Sequence 270, App

#### ALIGNMENTS

#### RESULT 1

US-10-032-221B-38  
; Sequence 38, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
; FILE REFERENCE: 2312/2082B (Formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 38  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T7-mutant (amino acids 73-97 of SEQ ID NO:10; methionine has be  
; OTHER INFORMATION: substituted for the leucine residue at position 77 of the full-  
; OTHER INFORMATION: ensth Tmstatatin molecule, and isoleucine has been substituted f  
; OTHER INFORMATION: valine at position 81, and asparagine has been substituted for  
; OTHER INFORMATION: spartic acid at position 83)  
US-10-032-221B-38

Mon Apr 5 07:53:10 2004

Query Match 100.0%; Score 148; DB 14; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
 DB 1 TMEFMCNINNVCFASRNDYSYWL 25

RESULT 2  
 US-10-206-699-306  
 ; Sequence 306, Application US/10206699  
 ; Publication No. US20030100510A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sundaramoorthy, M.  
 ; APPLICANT: Hudson, B.  
 ; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
 ; FILE REFERENCE: MBHB 01-1017  
 ; CURRENT APPLICATION NUMBER: US/10/206,699  
 ; CURRENT FILING DATE: 2002-07-26  
 ; PRIOR APPLICATION NUMBER: US 60/308,523  
 ; PRIOR FILING DATE: 2001-07-27  
 ; PRIOR APPLICATION NUMBER: US 60/351,289  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: US 60/366,854  
 ; PRIOR FILING DATE: 2002-03-22  
 ; PRIOR APPLICATION NUMBER: US 60/385,362  
 ; PRIOR FILING DATE: 2002-06-03  
 ; NUMBER OF SEQ ID NOS: 307  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 306  
 ; LENGTH: 229  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: alpha 5 chain  
 US-10-206-699-306

Query Match 100.0%; Score 148; DB 14; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
 DB 59 TMEFMCNINNVCFASRNDYSYWL 83

RESULT 3  
 US-09-925-297-496  
 ; Sequence 496, Application US/09925297  
 ; Patent No. US20020081659A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA105  
 ; CURRENT APPLICATION NUMBER: US/09/925,297  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05989  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 928  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 496  
 ; LENGTH: 309  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (247)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-925-297-496

Query Match 100.0%; Score 148; DB 9; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-13;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
 DB 139 TMEFMCNINNVCFASRNDYSYWL 163

RESULT 4  
 US-10-206-699-302  
 ; Sequence 302, Application US/10206699  
 ; Publication No. US20030100510A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sundaramoorthy, M.  
 ; APPLICANT: Hudson, B.  
 ; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
 ; FILE REFERENCE: MBHB 01-1017  
 ; CURRENT APPLICATION NUMBER: US/10/206,699  
 ; CURRENT FILING DATE: 2002-07-26  
 ; PRIOR APPLICATION NUMBER: US 60/308,523  
 ; PRIOR FILING DATE: 2001-07-27  
 ; PRIOR APPLICATION NUMBER: US 60/351,289  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: US 60/366,854  
 ; PRIOR FILING DATE: 2002-03-22  
 ; PRIOR APPLICATION NUMBER: US 60/385,362  
 ; PRIOR FILING DATE: 2002-06-03  
 ; NUMBER OF SEQ ID NOS: 307  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 302  
 ; LENGTH: 229  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: alpha 1 chain  
 US-10-206-699-302

Query Match 98.0%; Score 145; DB 14; Length 229;  
 Best Local Similarity 96.0%; Pred. No. 5e-13;  
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMEFMCNINNVCFASRNDYSYWL 25  
 DB 59 TMEFMCNINNVCFASRNDYSYWL 83

RESULT 5  
 US-10-032-221B-2  
 ; Sequence 2, Application US/10032221B  
 ; Publication No. US2003014481A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kalluri, Raghuram  
 ; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
 ; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
 ; CURRENT APPLICATION NUMBER: US/10/032,221B  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00565  
 ; PRIOR FILING DATE: 2001-01-08  
 ; PRIOR APPLICATION NUMBER: US 09/625,191  
 ; PRIOR FILING DATE: 2000-07-21  
 ; PRIOR APPLICATION NUMBER: US 09/543,371  
 ; PRIOR FILING DATE: 2000-04-04  
 ; PRIOR APPLICATION NUMBER: US 09/479,118  
 ; PRIOR FILING DATE: 2000-01-07  
 ; PRIOR APPLICATION NUMBER: US 09/335,224  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: US 60/126,175  
 ; PRIOR FILING DATE: 1999-03-25  
 ; PRIOR APPLICATION NUMBER: US 60/089,689  
 ; PRIOR FILING DATE: 1998-06-17

; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-2

Query Match 98.0%; Score 145; DB 14; Length 229;  
Best Local Similarity 96.0%; Pred. No. 5e-13;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25  
Db 59 TWPFFMFCNNVNCNPFASRNDYSYWL 83

## RESULT 6

US-09-925-302-507  
; Sequence 507, Application US/09925302  
; Patent No. US2002004941A1  
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US/09/925,302  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 507  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (71)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-507

Query Match 98.0%; Score 145; DB 9; Length 406;  
Best Local Similarity 96.0%; Pred. No. 8.7e-13;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25  
Db 236 TWPFFMFCNNVNCNPFASRNDYSYWL 260

## RESULT 7

US-10-372-683-8  
; Sequence 8, Application US/10372683  
; Publication No. US20040009171A1  
; GENERAL INFORMATION:

; APPLICANT: GERRITSEN, MARY E.  
; APPLICANT: PEALE JR., FRANKLIN V.  
; APPLICANT: WU, THOMAS D.  
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA  
; FILE REFERENCE: P1928R1P1  
; CURRENT APPLICATION NUMBER: US/10/372,683  
; CURRENT FILING DATE: 2003-02-21  
; PRIOR APPLICATION NUMBER: US 10/271,690  
; PRIOR FILING DATE: 2002-10-16  
; PRIOR APPLICATION NUMBER: US 60/344,534  
; PRIOR FILING DATE: 2001-10-18  
; NUMBER OF SEQ ID NOS: 49  
; SEQ ID NO 8  
; LENGTH: 1669  
; TYPE: PRT  
; ORGANISM: Homo sapien

## US-10-372-683-8

Query Match 98.0%; Score 145; DB 15; Length 1669;  
Best Local Similarity 96.0%; Pred. No. 3.4e-12;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25  
Db 1499 TWPFFMFCNNVNCNPFASRNDYSYWL 1523

## RESULT 8

US-10-032-221B-37  
; Sequence 37, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B

; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 37  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T7 (amino acids 73-97 of SEQ ID NO:10)  
US-10-032-221B-37

Query Match 93.9%; Score 139; DB 14; Length 25;  
Best Local Similarity 88.0%; Pred. No. 4.4e-13;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25  
Db 1 TWPFFMFCNNVNCNPFASRNDYSYWL 25

## RESULT 9

US-10-032-221B-26  
; Sequence 26, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B

; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224

PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 26  
LENGTH: 79  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)  
US-10-032-221B-26

Query Match 93.9%; Score 139; DB 14; Length 79;  
Best Local Similarity 88.0%; Pred. No. 1.3e-12;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFSRNDYSYWL 25  
|||:||||:|||||  
DB 20 TMPFLFCNVNVCNFSRNDYSYWL 44

## RESULT 10

US-10-032-221B-33  
Sequence 33, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 33  
LENGTH: 88  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)  
US-10-032-221B-33

Query Match 93.9%; Score 139; DB 14; Length 88;  
Best Local Similarity 88.0%; Pred. No. 1.5e-12;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFSRNDYSYWL 25  
|||:||||:|||||  
DB 29 TMPFLFCNVNVCNFSRNDYSYWL 53

## RESULT 11

US-10-032-221B-34  
Sequence 34, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram

TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 34  
LENGTH: 88  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine has been substituted for the cysteine residue at position 125 of the full-length Tumstatin molecule)  
OTHER INFORMATION: the full-length Tumstatin molecule)  
US-10-032-221B-34

Query Match 93.9%; Score 139; DB 14; Length 88;  
Best Local Similarity 88.0%; Pred. No. 1.5e-12;  
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMPFMFCNNVNCNFSRNDYSYWL 25  
|||:||||:|||||  
DB 29 TMPFLFCNVNVCNFSRNDYSYWL 53

## RESULT 12

US-10-032-221B-20  
Sequence 20, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 20  
LENGTH: 124  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)  
US-10-032-221B-20





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 5.87167 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221B-38

Perfect score: 148  
Sequence: 1 TWPFFMFCNNVNCNFASRNDYSYWL 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	264	4	US-09-589-927-10
2	148	100.0	264	4	US-09-277-665-10
3	148	100.0	264	4	US-09-589-987-10
4	145	98.0	260	4	US-09-589-927-2
5	145	98.0	260	4	US-09-277-665-2
6	145	98.0	260	4	US-09-589-987-2
7	140	94.6	471	2	US-08-399-889-24
8	140	94.6	471	3	US-09-157-364-24
9	140	94.6	471	3	US-09-439-997-2
10	139	93.9	211	4	US-09-512-563C-46
11	139	93.9	218	3	US-08-399-889-25
12	139	93.9	218	3	US-09-167-364-25
13	139	93.9	218	3	US-09-439-997-4
14	139	93.9	268	4	US-09-589-927-6
15	139	93.9	268	4	US-09-277-665-6
16	139	93.9	268	4	US-09-589-987-6
17	113	76.4	260	4	US-09-589-927-12
18	113	76.4	260	4	US-09-277-665-12
19	113	76.4	260	4	US-09-589-987-12
20	104	70.3	258	4	US-09-589-927-4
21	104	70.3	258	4	US-09-277-665-4
22	104	70.3	258	4	US-09-589-987-4
23	104	70.3	260	4	US-09-589-927-8
24	104	70.3	260	4	US-09-277-665-8
25	104	70.3	260	4	US-09-589-987-8
26	89	60.1	1694	1	US-08-494-168-2
27	46	31.1	83	4	US-09-543-681A-8079

28 46 31.1 160 4 US-09-589-927-10 Sequence 7, Appli  
29 45.5 30.7 704 3 US-08-752-832A-2 Sequence 2, Appli  
30 45 30.4 867 4 US-09-668-673B-2 Sequence 2, Appli  
31 45 30.4 871 4 US-09-773-426A-3 Sequence 3, Appli  
32 45 30.4 1611 4 US-09-668-673B-16 Sequence 16, Appli  
33 44 29.7 49 1 US-07-865-166A-6 Sequence 6, Appli  
34 44 29.7 359 1 US-08-307-382-2 Sequence 2, Appli  
35 44 29.7 359 1 US-08-366-779-2 Sequence 2, Appli  
36 44 29.7 359 1 US-08-478-727-2 Sequence 2, Appli  
37 44 29.7 359 1 US-08-473-508-2 Sequence 2, Appli  
38 44 29.7 359 1 US-08-789-936-2 Sequence 2, Appli  
39 44 29.7 359 2 US-08-833-610-6 Sequence 6, Appli  
40 44 29.7 359 3 US-08-834-033A-16 Sequence 16, Appli  
41 44 29.7 359 4 US-08-934-254-2 Sequence 2, Appli  
42 44 29.7 359 4 US-09-377-452-6 Sequence 6, Appli  
43 44 29.7 359 4 US-09-685-775-2 Sequence 2, Appli  
44 44 29.7 575 4 US-09-107-532A-4554 Sequence 4554, Ap  
45 43.5 29.4 354 4 US-09-574-942-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-589-927-10  
; Sequence 10, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589.927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-10

Query Match 100.0%; Score 148; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 3.7e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFFMFCNNVNCNFASRNDYSYWL 25  
Db 94 TWPFFMFCNNVNCNFASRNDYSYWL 118

RESULT 2  
US-09-277-665-10  
; Sequence 10, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277.665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-10

Query Match 100.0%; Score 148; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 3.7e-13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 94 TMPFMFCNNVNCNFASRNDYSYWL 118

RESULT 3
US-09-589-987-10
; Sequence 10, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-10

Query Match 100.0%; Score 148; DB 4; Length 264;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 94 TMPFMFCNNVNCNFASRNDYSYWL 118

RESULT 4
US-09-589-927-2
; Sequence 2, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-2

Query Match 98.0%; Score 145; DB 4; Length 260;
Best Local Similarity 96.0%; Pred. No. 9.6e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 90 TMPFMFCNNVNCNFASRNDYSYWL 114

RESULT 5
US-09-277-665-2
; Sequence 2, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26

Query Match 94.6%; Score 140; DB 2; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 90 TMPFMFCNNVNCNFASRNDYSYWL 114

RESULT 6
US-09-589-987-2
; Sequence 2, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-2

Query Match 98.0%; Score 145; DB 4; Length 260;
Best Local Similarity 96.0%; Pred. No. 9.6e-13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 90 TMPFMFCNNVNCNFASRNDYSYWL 114

RESULT 7
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match 94.6%; Score 140; DB 2; Length 471;
Best Local Similarity 92.0%; Pred. No. 8.8e-12;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TMPFMFCNNVNCNFASRNDYSYWL 25
Db 90 TMPFMFCNNVNCNFASRNDYSYWL 114
```

Db 300 TWPFLFCNINDVCFASRNDYSYWL 324

RESULT 8

US-09-167-364-24

Sequence 24, Application US/09167364

Patent No. 6007980

GENERAL INFORMATION:

APPLICANT: Reeder, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263B

CURRENT APPLICATION NUMBER: US/09/167,364

CURRENT FILING DATE: 1998-10-07

EARLIER APPLICATION NUMBER: 08/399889

EARLIER FILING DATE: 1995-03-07

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24

LENGTH: 471

TYPE: PRT

ORGANISM: Calf

US-09-167-364-24

Query Match 94.6%; Score 140; DB 3; Length 471;

Best Local Similarity 92.0%; Pred. No. 8.8e-12;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFLFCNINDVCFASRNDYSYWL 25

Db 300 TWPFLFCNINDVCFASRNDYSYWL 324

RESULT 9

US-09-439-897-2

Sequence 2, Application US/09439897

Patent No. 6277558

GENERAL INFORMATION:

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 95-1263-C

CURRENT APPLICATION NUMBER: US/09/439,897

CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 65

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 471

TYPE: PRT

ORGANISM: Bos taurus

US-09-439-897-2

Query Match 94.6%; Score 140; DB 3; Length 471;

Best Local Similarity 92.0%; Pred. No. 8.8e-12;

Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFLFCNINDVCFASRNDYSYWL 25

Db 300 TWPFLFCNINDVCFASRNDYSYWL 324

RESULT 10

US-09-512-563C-46

Sequence 46, Application US/09512563C

Patent No. 6579969

GENERAL INFORMATION:

APPLICANT: Saus, Juan

TITLE OF INVENTION: Goodpasture Binding Protein

FILE REFERENCE: 98-723-A

CURRENT APPLICATION NUMBER: US/09/512,563C

CURRENT FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/121,483

PRIOR FILING DATE: 1999-02-24

NUMBER OF SEQ ID NOS: 63

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 46

LENGTH: 211

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GPDV

US-09-512-563C-46

Query Match 93.9%; Score 139; DB 4; Length 211;

Best Local Similarity 88.0%; Pred. No. 5.3e-12;

Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFLFCNINDVCFASRNDYSYWL 25

Db 73 TWPFLFCNINDVCFASRNDYSYWL 97

RESULT 11

US-08-399-889-25

Sequence 25, Application US/08399889B

Patent No. 5973120

GENERAL INFORMATION:

APPLICANT: Reeder, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263A

CURRENT APPLICATION NUMBER: US/08/399,889B

CURRENT FILING DATE: 1995-03-07

EARLIER APPLICATION NUMBER: 07/621091

EARLIER FILING DATE: 1990-11-30

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 25

LENGTH: 218

TYPE: PRT

ORGANISM: Human

US-08-399-889-25

Query Match 93.9%; Score 139; DB 2; Length 218;

Best Local Similarity 88.0%; Pred. No. 5.4e-12;

Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWPFLFCNINDVCFASRNDYSYWL 25

Db 47 TWPFLFCNINDVCFASRNDYSYWL 71

RESULT 12

US-09-167-364-25

Sequence 25, Application US/09167364

Patent No. 6007980

GENERAL INFORMATION:

APPLICANT: Reeder, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263B

CURRENT APPLICATION NUMBER: US/09/167,364

CURRENT FILING DATE: 1998-10-07

EARLIER APPLICATION NUMBER: 08/399889

EARLIER FILING DATE: 1995-03-07

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 25

LENGTH: 218

TYPE: PRT

ORGANISM: Human

US-09-167-364-25

Query Match 93.9%; Score 139; DB 3; Length 218;

Best Local Similarity 88.0%; Pred. No. 5.4e-12; Indels 0; Gaps 0;  
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25  
Db 47 TMPFLFCNVNVCNFAFRNDYSYWL 71

## RESULT 13

US-09-439-897-4  
; Sequence 4, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1283-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-439-897-4

Query Match 93.9%; Score 139; DB 3; Length 218;  
Best Local Similarity 88.0%; Pred. No. 5.4e-12; Indels 0; Gaps 0;  
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25  
Db 47 TMPFLFCNVNVCNFAFRNDYSYWL 71

## RESULT 14

US-09-589-927-6  
; Sequence 6, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 268  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-6

Query Match 93.9%; Score 139; DB 4; Length 268;  
Best Local Similarity 88.0%; Pred. No. 6.7e-12; Indels 0; Gaps 0;  
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25  
Db 97 TMPFLFCNVNVCNFAFRNDYSYWL 121

## RESULT 15

US-09-277-665-6  
; Sequence 6, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 268  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-6

Query Match 93.9%; Score 139; DB 4; Length 268;  
Best Local Similarity 88.0%; Pred. No. 6.7e-12; Indels 0; Gaps 0;  
Matches 22; Conservative 3; Mismatches 0;

QY 1 TMPFMFCINNVNCFASRNDYSYWL 25  
Db 97 TMPFLFCNVNVCNFAFRNDYSYWL 121

Search completed: April 5, 2004, 07:07:25  
Job time : 6.87167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	96.7	220	2 B49736	collagen alpha 3(I
2	146	96.7	1670	1 CGHU3B	collagen alpha 3(I
3	145	95.0	471	2 A39024	collagen alpha 3(I
4	140	92.7	161	2 S49488	collagen alpha 3(I
5	140	92.7	246	2 I48302	collagen alpha 3(I
6	130	85.1	253	2 I48304	collagen alpha 3(I
7	130	85.1	754	2 A5267	collagen alpha 5(I
8	130	85.1	1691	1 S22917	collagen alpha 5(I
9	129	85.4	258	2 B61228	collagen alpha 1(I
10	129	85.4	1669	1 CGHU4B	collagen alpha 1(I
11	129	85.4	1669	1 CGMS4B	collagen alpha 1(I
12	120	79.5	1747	2 A54121	collagen alpha-4 c
13	120	79.5	1752	2 A45407	collagen alpha 3(I
14	115	76.2	1758	2 T29350	hypothetical prote
15	115	76.2	1759	2 T29351	collagen alpha 2(I
16	115	76.2	1763	2 S16366	collagen alpha 2(I
17	111	73.5	261	2 A34476	collagen alpha 2(I
18	107	70.9	1744	2 S40991	collagen alpha 1(I
19	105	69.5	1691	1 CGHU6B	collagen alpha 6(I
20	102	67.5	1775	2 A61228	collagen alpha 2(I
21	102	67.5	1707	2 A33526	collagen alpha 2(I
22	102	67.5	1712	1 CGHU2B	collagen type IV a
23	89	58.9	1761	2 T13990	collagen alpha 4(I
24	88	58.3	312	2 I48303	collagen alpha 4(I
25	88	58.3	623	2 A45137	collagen alpha 4(I
26	88	58.3	1690	1 CGHU1B	collagen alpha 4(I
27	88	58.3	1775	2 A31893	collagen alpha 1(I
28	87	57.6	453	2 S18804	collagen alpha 4(I
29	62	41.1	79	2 C43928	probable collagen

30 50 33.1 489 2 B42815  
31 49.5 32.8 419 2 S41607  
32 49 32.5 334 2 S04324  
33 48 31.8 257 2 H5419  
34 47 31.1 363 2 T37630  
35 46 30.5 98 1 F2NTR  
36 45.5 30.1 397 2 F97707  
37 45.5 30.1 610 2 C70126  
38 45.5 30.1 1743 2 T18279  
39 45 29.8 245 2 D72417  
40 45 29.8 963 2 S45167  
41 45 29.8 1095 2 PC1114  
42 44.5 29.5 504 2 S24314  
43 44 29.1 205 2 E82601  
44 44 29.1 221 2 D84516  
45 44 29.1 427 2 T14421

DNA primase Sog210  
atrolysin A (EC 3.  
hypothetical prote  
hypothetical prote  
protein disulfide-  
photosystem II pro  
acyl-CoA desaturas  
DNA mismatch repai  
multidrug resistanc  
conserved hypothet  
chitin synthase (E  
SKDCD35 protein -  
bacterial leucyl a  
conserved hypothet  
hypothetical prote  
S-locus-specific g

#### ALIGNMENTS

##### RESULT 1

B49736  
collagen alpha 3(IV) chain, medium splice form - human (fragment)  
N;Contains: collagen alpha 3(IV) chain, splice form GP-V  
C;Species: Homo sapiens (man)  
C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text\_change 17-Mar-2000  
C;Accession: B49736; D49736; S69111  
R;Peng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A;Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.  
A;Reference number: A49736; MUID:94124597; PMID:8294492  
A;Accession: B49736  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 159-220 <FEN1>  
A;Accession: D49736  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: mRNA  
A;Residues: 22-220 <FEN2>  
A;Cross-references: GB:U02519; NID:9409106; PIDN:AAA18942.1; PID:9409107  
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank  
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Carvera, J.; Wi  
Eur. J. Biochem. 229, 754-760, 1995  
A;Title: Characterization and expression of multiple alternatively spliced transcripts  
utocantigen and one of its alternative forms.  
A;Reference number: S69111; MUID:95278230; PMID:7758473  
A;Accession: S69111  
A;Molecule type: mRNA  
A;Residues: 1-45,169-204,'L',206-220 <PEN>  
A;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace  
F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pred  
F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status  
F;22-220/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 96.7%; Score 146; DB 2; Length 220;

Best Local Similarity 100.0%; Pred. No. 4.7e-14; Mismatches 0; Gaps 0;  
Matches 26; Conservative 0; Indels 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

DB 78 QRTTTPFLFCNVNDVCFASRNDYS 103

##### RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human  
N;Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
C;Species: Homo sapiens (man)  
C;Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
C;Accession: A54763; A43928; A44043; A45971; A39786  
R;Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
A;Reference number: A54763; MUID:94364994; PMID:8083201  
A;Accession: A54763  
A;Molecule type: mRNA  
A;Residues: 1-1670 <VAR>  
A;Cross-references: GB:X80031; NID:G577563; PID:G577564  
A;Experimental source: kidney  
R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha  
A;Reference number: A43928; MUID:92147878; PMID:1737849  
A;Accession: A43928  
A;Molecule type: mRNA  
A;Residues: 1331-1524, 'I', 1526-1670 <TUR>  
A;Cross-references: GB:M81379  
A;Experimental source: kidney  
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 19780-19784, 1992  
A;Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture  
A;Reference number: A44043; MUID:93015826; PMID:1400291  
A;Accession: A44043  
A;Molecule type: DNA; mRNA  
A;Residues: 1386-1670 <QUI>  
A;Cross-references: GB:M92993; NID:G177895; PIDN:AA21610.1; PID:G177896  
A;Note: sequence extracted from NCBI backbone (NCBIP:115597)  
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 269, 17358, 1994  
A;Reference number: A44739; MUID:94274734; PMID:8006044  
A;Contents: annotation; erratum; correction to intronic sequence in A44043  
R;Bernal, D.; Quinones, S.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A;Reference number: A45971; MUID:93280184; PMID:8505332  
A;Accession: A45971  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1427-1444 <BER>  
A;Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly ident  
R;Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Reiders, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1997  
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
A;Reference number: A39786; MUID:91353570; PMID:1682840  
A;Accession: A39786  
A;Molecule type: mRNA  
A;Residues: 1453-1593, 'A', 1595-1670 <MOR>  
A;Cross-references: GB:S55790; NID:G234418; PIDN:AA819637.1; PID:G234419  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-q37  
A;Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1595/3; 1643/2 #status incomplete  
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
er associations in the interrupted helical domain (with disulfide and desmosine cross-l  
C;Function:  
A;Description: minor structural component of extracellular basement membrane in kidney g  
A;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
F;1-238/Domain: signal sequence #status predicted <STG>  
F;29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <VAT>  
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>

F;43-1438/Region: interrupted helical  
F;791-793/Region: cell attachment (R-G-D) motif  
F;996-998/Region: cell attachment (R-G-D) motif  
F;1154-1156/Region: cell attachment (R-G-D) motif  
F;1306-1308/Region: cell attachment (R-G-D) motif  
F;1345-1347/Region: cell attachment (R-G-D) motif  
F;1432-1434/Region: cell attachment (R-G-D) motif  
F;1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F;31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred  
F;253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F;1505-1511.1616-1622/Disulfide bonds: #status predicted  
F;1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
Query Match 96.7%; Score 146; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 3.3e-13;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
DB 1495 QRFTHMPFLFCNVNDVCFASRNDYS 1520  
RESULT 3  
A39024  
collagen alpha 3(IV) chain - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text change 13-Aug-1999  
C;Accession: A39024; S20672; S17802; A35167; S33747; S20815  
R;Morrison, K.E.; Germino, G.G.; Reiders, S.T.  
J. Biol. Chem. 266, 34-39, 1991  
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
A;Reference number: A39024; MUID:91093146; PMID:1985905  
A;Accession: A39024  
A;Molecule type: mRNA  
A;Residues: 1-471 <NOR>  
A;Cross-references: EMBL:M63139; NID:G162886; PIDN:AA62708.1; PID:G162887  
R;Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
A;Reference number: S18432; MUID:87222419; PMID:2438283  
A;Accession: S20672  
A;Molecule type: protein  
A;Residues: 227-228, 'X', 230-244 <BUT>  
R;Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
J. Biol. Chem. 263, 13374-13380, 1988  
A;Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
A;Reference number: S17802; MUID:88330844; PMID:3417661  
A;Accession: S17802  
A;Molecule type: protein  
A;Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5466-5469, 1990  
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
A;Reference number: A35167; MUID:90202779; PMID:2318822  
A;Accession: A35167  
A;Molecule type: protein  
A;Residues: 236-258 <GUN>  
R;Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No  
J. Biol. Chem. 266, 15318-15324, 1991  
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol  
A;Reference number: A39419; MUID:91332055; PMID:1869555  
A;Accession: C39419  
A;Molecule type: protein  
A;Residues: 236-255 <GU2>  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;  
F;1-238/Domain: collagenous (fragment) #status predicted <COL>  
F;239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F;239-353/Domain: repeat NC1 #status predicted <NC11>  
F;354-471/Domain: repeat NC1 #status predicted <NC12>



F:232,238/Modified site: hydroxyproline (Pro) #status experimental  
F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 96.0%; Score 145; DB 2; Length 471;  
Best Local Similarity 96.2%; Pred. No. 1.4e-13;  
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVNCVFASRNDYS 27  
DB 236 QRFTHMPFLFCNVNDVNCVFASRNDYS 321

## RESULT 4

S49488  
collagen alpha 3(IV) chain - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: S49488  
R:Oberbaumer, I.

submitted to the EMBL Data Library, October 1994  
A:Description: Cloning of the NC1 domains fo the minor collagen IV chains of mouse via B  
ells

A:Reference number: S49487

A:Accession: S49488

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-161 <OBS>

A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G5559916

C:Superfamily: collagen alpha 1(IV) chain

Query Match 92.7%; Score 140; DB 2; Length 161;

Best Local Similarity 92.3%; Pred. No. 2.7e-13;

Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVNCVFASRNDYS 27  
DB 4 QRFTHMPFLFCNVNDVNCVFASRNDYS 29

## RESULT 5

I48302  
collagen alpha 3(IV) chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
C:Accession: I48302; S47278  
R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48302

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-246 <RES>

A:Cross-references: EMBL:X35166; NID:G535197; PID:G535198

C:Superfamily: collagen alpha 1(IV) chain

Query Match 92.7%; Score 140; DB 2; Length 246;

Best Local Similarity 92.3%; Pred. No. 4e-13;

Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVNCVFASRNDYS 27  
DB 71 QRFTHMPFLFCNVNDVNCVFASRNDYS 96

## RESULT 6

I48304  
collagen alpha 5(IV) chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999  
C:Accession: I48304; S47280  
R:Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
A:Reference number: A54979; MUID:95050957; PMID:7962065

A:Accession: I48304

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <RES>

A:Cross-references: EMBL:X35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C:Superfamily: collagen alpha 1(IV) chain

Query Match 86.1%; Score 130; DB 2; Length 253;

Best Local Similarity 80.8%; Pred. No. 1.2e-11;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVNCVFASRNDYS 27  
DB 79 RRFSTMPFMCNINNVNCFASRNDYS 104

## RESULT 7

A55267  
collagen alpha 5(IV) chain - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: A55267

R:Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-1

en type IV

A:Reference number: A55267; MUID:94224868; PMID:8171024

A:Accession: A55267

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-754 <ZHE>

A:Cross-references: GB:U07888; NID:G469547; PIDN:AB60258.1; PID:G469548

C:Superfamily: collagen alpha 1(IV) chain

Query Match 86.1%; Score 130; DB 2; Length 754;

Best Local Similarity 80.8%; Pred. No. 3.4e-11;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVNCVFASRNDYS 27  
DB 587 RRFSTMPFMCNINNVNCFASRNDYS 612

## RESULT 8

S22917  
collagen alpha 5(IV) chain precursor, renal splice form - human  
N:Alternate names: procollagen alpha 5(IV) chain  
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence\_revision 27-Feb-1997 #text\_change 21-Jul-2000

C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A3

R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.

J. Biol. Chem. 267, 12475-12481, 1992

A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident

n Alport syndrome patient.

A:Reference number: S22917; MUID:92316923; PMID:1352287

A:Accession: S22917

A:Molecule type: mRNA

A:Residues: 1-967 <ZHO>

A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234

R:Zhou, J.; Leinonen, A.; Tryggvason, K.

J. Biol. Chem. 269, 6608-6614, 1994

A:Title: Structure of the human type IV collagen COL4A5 gene.

A:Reference number: A54365; MUID:94165049; PMID:8120014

A:Accession: A54365

A:Molecule type: DNA

A:Residues: 1-922 <ZH2>

A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAAC27816.1; PI

R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paape, A.; Tryggva

Science 261, 1167-1169, 1993

A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited s

[illegible]

Db 84 RKFTMPFLFCNNVNCVFNDRYS 109

## RESULT 10

CGHU4B

collagen alpha 1(IV) chain precursor - human  
N;Alternate names: procollagen alpha 1(IV) chain

C;Species: Homo sapiens (man)  
C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58  
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989

A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll  
A;Reference number: S16876; MUID:89340433; PMID:2701944

A;Accession: S16876

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1669 <SO11>

A;Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAAS3098.1; PID:G180803

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988

R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are

A;Reference number: A92690; MUID:89034231; PMID:3182844

A;Accession: A32117

A;Molecule type: DNA

A;Residues: 1-28 <SO12>

A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PID:G953233

R;Poeschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c

A;Reference number: S02738; MUID:89030632; PMID:2846280

A;Accession: S02738

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-6, 'L', '8-28 <POE>

A;Cross-references: EMBL:X12784; NID:G30072

R;Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.

Eur. J. Biochem. 168, 529-536, 1987

A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem

A;Reference number: S00048; MUID:88029471; PMID:3311751

A;Accession: S00048

A;Molecule type: mRNA

A;Residues: 1-318, 'A', 320-944 <BRA1>

A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAAR29075.1; PID:G30067

A;Accession: S25826

A;Molecule type: protein

A;Residues: 271-318, 'A', 320-554 <BRA2>

R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S

A;Reference number: A23115; MUID:86004708; PMID:4043082

A;Accession: A23115

A;Molecule type: protein

A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A;Experimental source: placenta

A;Note: the amino end of the mature form is blocked

R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.

FEBS Lett. 225, 188-194, 1987

A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty

A;Reference number: S00207; MUID:88083584; PMID:3691802

A;Accession: S00207

A;Molecule type: mRNA

A;Residues: 244-530 <SO13>

A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAAG68698.1; PID:G29549

R;Eble, J.A.; Golbig, R.; Mann, K.; Kuehn, K.

EMBO J. 12, 4795-4802, 1993

A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen

A;Reference number: S39614; MUID:94038963; PMID:8223488

A;Accession: S39614

A;Molecule type: protein

A;Residues: 371-554 <EBL>

R;Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid  
A;Reference number: A02863; MUID:85003629; PMID:6434307

A;Accession: A02863

A;Molecule type: protein

A;Residues: 534-718, 'D', 720-835, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99

A;Experimental source: placenta

R;Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A;Title: Pepsin fragments of human placental basement-membrane collagens showing inter  
A;Reference number: S16908; MUID:82005835; PMID:6792033

A;Accession: A58517

A;Molecule type: protein

A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-  
R;MacWhirter, R.S.; Benson, V.A.; Lovelillo, K.T.; van der Rest, M.; Fietzek, P.P.

Biochemistry 22, 4940-4948, 1983

A;Title: Isolation and characterization of pepsin-solubilized human basement membrane  
A;Reference number: S16910; MUID:84053346; PMID:6416291

A;Accession: S16910

A;Molecule type: protein

A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 94

A;Experimental source: placenta

R;Pillalajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;

J. Biol. Chem. 260, 7681-7687, 1985

A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen

A;Reference number: S01466; MUID:85207819; PMID:2581969

A;Accession: S01466

A;Molecule type: mRNA

A;Residues: 1256-1669 <PIH>

A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PID:G180424  
R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kafalides, N.A.

Proc. Natl. Acad. Sci. U.S.A. 82, 3648-3653, 1985

A;Title: Restricted homology between human alpha-1 type IV and other procollagen chain  
A;Reference number: S16879; MUID:85216555; PMID:2582422

A;Accession: S16879

A;Molecule type: mRNA

A;Residues: 1259-1669 <BRI>

A;Cross-references: EMBL:M11315; NID:G180817; PIDN:AAAS2042.1; PID:G180818

R;Oberbauer, I.; Laurent, M.; Schwaz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss

Eur. J. Biochem. 147, 217-224, 1985

A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha  
A;Reference number: A02864; MUID:85127033; PMID:12578961

A;Accession: S19091

A;Molecule type: protein

A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X'

R;Siebold, B.; Deutzmann, R.; Kuehn, K.

Eur. J. Biochem. 176, 617-624, 1988

A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxy  
A;Reference number: S02550; MUID:89005112; PMID:2844531

A;Contents: annotation; disulfide bonds

C;Genetics:

A;Gene: GDB:COL4A1

A;Cross-references: GDB:119791; OMIM:120130

A;Map position: 13q34-13q34

A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 23  
/; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 1

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha

C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha





T29351  
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans  
N:Alternate names: collagen alpha 2(IV) chain precursor c1b-1  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T29351  
R:Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid F01G12.  
A:Reference number: 220811  
A:Accession: T29351  
A>Status: preliminary; translated from CB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1759 <WUX>  
A:Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a  
A:Experimental source: strain Bristol N2; clone F01G12  
C:Genetics:  
A:Gene: CESP:F01G12.5a  
A:Map position: X  
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 76.2%; Score 115; DB 2; Length 1759;  
Best Local Similarity 76.9%; Pred. NO. 1.2e-08;  
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFSTMPFLFCNVNDVCFASRNDYS 27  
|||:|||||:|:|||||  
Db 1582 QRFSTMPFLFCDFNNVCNYSRNDKS 1607

Search completed: April 5, 2004, 07:05:37  
Job time : 6.5569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.39952 Seconds  
(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQFTTFMPLFCNVNDVCNFSARNDS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	96.7	1670	1 CA34 HUMAN	Q01955 homo sapien
2	145	96.0	471	1 CA34 BOVIN	Q28084 bos taurus
3	130	86.1	754	1 CA54 CANEA	Q28247 canis famill
4	129	86.1	1685	1 CA54 HUMAN	P29400 homo sapien
5	129	85.4	1669	1 CA14 HUMAN	P02462 homo sapien
6	129	85.4	1669	1 CA14 MOUSE	P02463 mus musculus
7	115	76.2	1763	1 CA24 ASCSU	P27393 ascaris suu
8	111	73.5	1758	1 CA24 CAEBL	P17140 caenorhabdi
9	107	70.9	1758	1 CA14 CAEBL	P17139 caenorhabdi
10	105	69.5	1691	1 CA64 HUMAN	Q14031 homo sapien
11	102	67.5	1707	1 CA24 MOUSE	P08122 mus musculus
12	102	67.5	1712	1 CA24 HUMAN	P08572 homo sapien
13	88	58.3	623	1 CA44 RABIT	P55787 cryptolagus
14	88	58.3	1690	1 CA44 HUMAN	P53420 homo sapien
15	88	58.3	1775	1 CA14 DROME	P08120 drosophila
16	87	57.6	453	1 CA44 BOVIN	Q29442 bos taurus
17	49	32.5	260	1 NPA HUMAN	Q31445 homo sapien
18	47.5	31.5	356	1 CMO_PIRFU	Q51741 pyrococcus
19	46	30.5	854	1 BAL HUMAN	Q01955 homo sapien
20	45.5	30.1	610	1 MUTL BORBU	Q01955 homo sapien
21	45.5	30.1	1743	1 TAGC DICDI	Q51229 borrelia bu
22	45	29.8	333	1 AMR1 HUMAN	Q23868 dictyosteli
23	45	29.8	344	1 AMR1 MOUSE	Q394x0 homo sapien
24	45	29.8	963	1 CUS2 YEAST	Q3jht5 mus muscu
25	45	29.8	1095	1 C25 SACKL	P4180 saccharomyc
26	44.5	29.5	504	1 AMPX VIBPR	Q02342 saccharomyc
27	43.5	28.8	334	1 Y092 RICPR	Q01693 vibrio prot
28	43.5	28.8	663	1 MM02 CHICK	Q02e55 rickettsia
29	43	28.5	256	1 YNCJ BACSU	Q30611 gallus gall
30	43	28.5	301	1 RC66 SCHPO	P39608 bacillus su
31	43	28.5	385	1 CHEB BORBU	Q94553 schizosacch
32	43	28.5	464	1 SYE2 COXBU	Q45047 borrelia bu
33	43	28.5	599	1 GP63 LEICH	Q83bl6 coxiella bu
					P15706 leishmania

## ALIGNMENTS

### RESULT 1

CA34\_HUMAN STANDARD; PRT; 1670 AA.  
ID CA34\_HUMAN STANDARD; PRT; 1670 AA.  
AC Q01955; Q9BOT2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE 10-OCT-2003 (Rel. 42; Last annotation update)  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=94364994; PubMed=8083201;  
RX Mariyama M., Leinonen A., Mochizuki T., Trygvason K., Readers S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues";  
RL J. Biol. Chem. 269:23013-23017(1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;  
GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND  
CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;  
PRO-574; GLU-1269 AND PRO-1474  
RX MEDLINE=21064696; PubMed=1134255;  
RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome";  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
RN [4]  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:15780-15784(1992).  
RN [5]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Readers S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen";  
RL Am. J. Hum. Genet. 49:545-554(1991).  
RN [6]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RX MEDLINE=92147878; PubMed=1737849;  
RC TISSUE=Kidney;  
RN

34 42.5 28.1 361 1 ALR CORGL  
35 42.5 28.1 407 1 NKIR\_HUMAN  
36 42.5 28.1 407 1 NKIR\_MOUSE  
37 42.5 28.1 407 1 NKIR\_RAT  
38 42 27.8 354 1 GBA2 SCHPO  
39 42 27.8 358 1 V242\_FOWPV  
40 42 27.8 505 1 CPDB\_MOUSE  
41 42 27.8 794 1 SUV5\_ARATH  
42 42 27.8 834 1 A1IM\_YEAST  
43 42 27.8 976 1 HMDH\_GIBFU  
44 42 27.8 1060 1 NKCL\_MANSE  
45 42 27.8 1260 1 ALS1\_CANAL

Q8tau9 corynebacte  
P25i03 homo sapien  
P30548 mus musculu  
P14600 rattus norv  
Q04665 schizosacch  
Q91428 fowlpox vir  
P24457 mus musculu  
O82175 arabidopsi  
P03875 saccharomyc  
Q12577 gibberella  
Q25479 marduca sex  
P46590 candida alb



RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D., cloning of the human Goodpasture antigen demonstrates it  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT to be the alpha 3 chain of type IV collagen.";  
RL J. Clin. Invest. 89:592-601(1992).  
RN [7]  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RA Ding J.;  
RN Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RL [8]  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions.";  
RL J. Biol. Chem. 269:2342-2348(1994).  
RN [9]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9337506;  
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Ninomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q36.";  
RL FEBS Lett. 424:11-16(1998).  
RN [10]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced.";  
RL J. Biol. Chem. 268:12090-12094(1993).  
RN [11]  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemlink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,  
RA Reenders S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome.";  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist. Isoforms differ in  
CC the C-terminal part of the NC1 domain;  
CC Name=1;  
CC IsoId=Q01955-1; Sequences=Displayed;  
CC Name=2; Synonyms=V;  
CC IsoId=Q01955-2; Sequences=VSP\_001170;  
CC Name=3; Synonyms=L5;  
CC IsoId=Q01955-3; Sequences=VSP\_001171;  
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
CC colocalized and present only in basement membranes of kidney, eye,  
CC cochlea, lung and brain.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
CC -!- DISEASE: Antibodies against the NC1 domain of alpha3(IV) mediate  
CC the autoimmune disease Goodpasture syndrome [MIM:233450], which is  
CC characterized by hematuria and pulmonary hemorrhage.  
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive  
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder  
CC characterized by progressive glomerulonephritis, renal failure,  
CC hematuria, ocular abnormalities and deafness. The recessive form  
CC occurs equally between males and females.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X80031; CAA56335.1; -;  
CC EMBL; AJ288487; CAC36101.1; JOINED.  
CC EMBL; AJ288488; CAC36101.1; JOINED.  
CC EMBL; AJ288489; CAC36101.1; JOINED.  
CC EMBL; AJ288490; CAC36101.1; JOINED.  
CC EMBL; AJ288491; CAC36101.1; JOINED.  
CC EMBL; AJ288492; CAC36101.1; JOINED.  
CC EMBL; AJ288493; CAC36101.1; JOINED.  
CC EMBL; AJ288494; CAC36101.1; JOINED.  
CC EMBL; AJ288495; CAC36101.1; JOINED.  
CC EMBL; AJ288496; CAC36101.1; JOINED.  
CC EMBL; AJ288497; CAC36101.1; JOINED.  
CC EMBL; AJ288498; CAC36101.1; JOINED.  
CC EMBL; AJ288499; CAC36101.1; JOINED.  
CC EMBL; AJ288500; CAC36101.1; JOINED.  
CC EMBL; AJ288501; CAC36101.1; JOINED.  
CC EMBL; AJ288502; CAC36101.1; JOINED.  
CC EMBL; AJ288503; CAC36101.1; JOINED.  
CC EMBL; AJ288504; CAC36101.1; JOINED.  
CC EMBL; AJ288505; CAC36101.1; JOINED.  
CC EMBL; AJ288506; CAC36101.1; JOINED.  
CC EMBL; AJ288507; CAC36101.1; JOINED.  
CC EMBL; AJ288508; CAC36101.1; JOINED.  
CC EMBL; AJ288509; CAC36101.1; JOINED.  
CC EMBL; AJ288510; CAC36101.1; JOINED.  
CC EMBL; AJ288511; CAC36101.1; JOINED.  
CC EMBL; AJ288512; CAC36101.1; JOINED.  
CC EMBL; AJ288513; CAC36101.1; JOINED.  
CC EMBL; AJ288514; CAC36101.1; JOINED.  
CC EMBL; AJ288515; CAC36101.1; JOINED.  
CC EMBL; AJ288516; CAC36101.1; JOINED.  
CC EMBL; AJ288517; CAC36101.1; JOINED.  
CC EMBL; AJ288518; CAC36101.1; JOINED.  
CC EMBL; AJ288519; CAC36101.1; JOINED.  
CC EMBL; AJ288520; CAC36101.1; JOINED.  
CC EMBL; AJ288521; CAC36101.1; JOINED.  
CC EMBL; AJ288522; CAC36101.1; JOINED.  
CC EMBL; AJ288523; CAC36101.1; JOINED.  
CC EMBL; AJ288524; CAC36101.1; JOINED.  
CC EMBL; AJ288525; CAC36101.1; JOINED.  
CC EMBL; AJ288526; CAC36101.1; JOINED.  
CC EMBL; AJ288527; CAC36101.1; JOINED.  
CC EMBL; AJ288528; CAC36101.1; JOINED.  
CC EMBL; AJ288529; CAC36101.1; JOINED.  
CC EMBL; AJ288530; CAC36101.1; JOINED.  
CC EMBL; AJ288531; CAC36101.1; JOINED.  
CC EMBL; AJ288532; CAC36101.1; JOINED.  
CC EMBL; AJ288533; CAC36101.1; JOINED.  
CC EMBL; AJ288534; CAC36101.1; JOINED.  
CC EMBL; AJ288535; CAC36101.1; JOINED.  
CC EMBL; AJ288536; CAC36101.1; JOINED.

```

Query Match          96.7%; Score 145; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27
    |||||
DB 1495 QRFTTTPFLFCNVNDVCFASRNDYS 1520
    |||||

RESULT 2
CA34 BOVIN
ID CA34 BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093145; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reeders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
    encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
    alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
    of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=97222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
    Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
    membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
    glomerular basement membranes (GBM), forming a 'chicken-wire'
    meshwork together with laminins, proteoglycans and entactin/
    nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
    alpha 6(IV), each of which can form a triple helix structure
    with 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
    domain (NC1) at their C-terminus, frequent interruptions of the
    G-X-Y repeats in the long central triple-helical domain (which may
    cause flexibility in the triple helix), and a short N-terminal
    triple-helical 7S domain.
CC -1- FTM: Prolines at the third position of the tripeptide repeating
    unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- FTM: Type IV collagens contain numerous cysteine residues which
    are involved in inter- and intramolecular disulfide bonding. 12 of
    these, located in the NC1 domain, are conserved in all known type
  
```

```

IV collagens.
-1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M63139; AAA62708.1; -.
DR PIR: A39024; A39024.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR004442; Procollagen4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 4.
DR ProDom: PD03923; ProcollagenC4; 1.
DR SMART: SMO0111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match          96.0%; Score 145; DB 1; Length 471;
Best Local Similarity 96.2%; Pred. No. 6.6e-14;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27
    |||||
DB 296 QRFTTTPFLFCNVNDVCFASRNDYS 321
    |||||

RESULT 3
CA54 CANFA
ID CA54 CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sankey; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
    human X-linked hereditary nephritis resulting from a single base
    mutation in the gene encoding the alpha 5 chain of collagen type
    IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -1- FUNCTION: Type IV collagen is the major structural component of
    glomerular basement membranes (GBM), forming a 'chicken-wire'
    meshwork together with laminins, proteoglycans and entactin/
    nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
  
```



```

RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of α5(IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
RN [19]
RV VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
RN [20]
RV VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzone G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
RN [21]
RV VARIANTS AS VAL-420; 456-PRO→-PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY→-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
RX MEDLINE=99063529; PubMed=9848783.
RA Martin P., Heskari N., Zhou J., Leinonen A., Tuomelius T., Hertz J.M.,
RA Springer J., Shows T.B., Atkin C.L., Stykarsdottir U., Neumann H.,
RA Springate J., Shows T.C., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
RN [22]
RV VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
RN [23]
RV VARIANT AS ARG-822.
RP
Query Match 86.1%; Score 130; DB 1; Length 1685;
Best Local Similarity 80.8%; Pred. No. 3.9e-11;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps
Qy 2 QRFTMPFLFCNVNDVCFASRNDYS 27
Db 1511 RFSTMFPMFCNNVCNFASRNDYS 1536
RESULT 5
ID CA14_HUMAN
ID CA14_HUMAN STANDARD; PRT; 1669 AA.
AC P02462;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(Iv) chain precursor.
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RV SEQUENCE FROM N.A.
RX MEDLINE=89340433; PubMed=2701944;
RA Soiminen R., Huokari M., Ganguly A., Prockop D.J., Tryggyvason K.;
RT "Structural organization of the gene for the alpha 1 chain of human
RT type IV collagen.";
```

J. Biol. Chem. 264:13565-13571 (1989).  
[2]  
SEQUENCE OF 46-1257 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=8803584; PubMed=3691802;  
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
RT "Complete primary structure of the alpha 1-chain of human basement  
membrane (type IV) collagen.";  
RL FBBS Lett. 225:188-194 (1987).  
RN  
[3]  
SEQUENCE OF 1-943 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88029471; PubMed=3311751;  
RA Brazel D., Oberbaumer I., Dieringer H., Babel W., Glanville R.W.,  
Deutzmann R., Kuehn K.;  
RT "Completion of the amino acid sequence of the alpha 1 chain of human  
basement membrane collagen (type IV) reveals 21 non-triplet  
interruptions located within the collagenous domain.";  
RL Eur. J. Biochem. 168:529-536 (1987).  
RN  
[4]  
SEQUENCE OF 28-243.  
RX MEDLINE=86004708; PubMed=4043082;  
RA Glanville R.W., Qian R.O., Siebold B., Risteli J., Kuehn K.;  
RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
region (VS domain) of the alpha 1 (IV) chain of human basement  
membrane collagen.";  
RL Eur. J. Biochem. 152:213-219 (1985).  
RN  
[5]  
SEQUENCE OF 534-1447.  
RX MEDLINE=85003629; PubMed=6434307;  
RA Babel W., Glanville R.W.;  
RT "Structure of human-basement-membrane (type IV) collagen. Complete  
amino-acid sequence of a 914-residue-long pepsin fragment from the  
alpha 1 (IV) chain.";  
RL Eur. J. Biochem. 143:545-556 (1984).  
RN  
[6]  
SEQUENCE OF 1256-1669 FROM N.A.  
RX MEDLINE=85207819; PubMed=2581969;  
RA Pihlajantemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
Cheung M.-C., Prockop D.J., Boyd C.D.;  
RT "cDNA clones coding for the pro-alpha 1 (IV) chain of human type IV  
procollagen reveal an unusual homology of amino acid sequences in two  
halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687 (1985).  
RN  
[7]  
SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653 (1985).  
RN  
[8]  
SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89034231; PubMed=3182844;  
RA Soininen R., Huotari M., Hosikka S.L., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
collagen are divergently encoded on opposite DNA strands and have an  
overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220 (1988).  
RN  
[9]  
SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
carboxyterminal, non-collagenous aggregation and cross-linking domain  
of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624 (1988).  
RN  
-!- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.  
CC

CC CC  
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
alpha 6(IV), each of which can form a triple helix structure  
with 2 other chains to generate type IV collagen network.  
CC  
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC  
-!- PWM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC  
-!- PWM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC  
-!- PWM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M26576; AAA53098.1; JOINED.  
CC EMBL; J04217; AAA53098.1; JOINED.  
CC EMBL; M26550; AAA53098.1; JOINED.  
CC EMBL; M26540; AAA53098.1; JOINED.  
CC EMBL; M26542; AAA53098.1; JOINED.  
CC EMBL; M26543; AAA53098.1; JOINED.  
CC EMBL; M26544; AAA53098.1; JOINED.  
CC EMBL; M26545; AAA53098.1; JOINED.  
CC EMBL; M26546; AAA53098.1; JOINED.  
CC EMBL; M26547; AAA53098.1; JOINED.  
CC EMBL; M26537; AAA53098.1; JOINED.  
CC EMBL; M26538; AAA53098.1; JOINED.  
CC EMBL; M26548; AAA53098.1; JOINED.  
CC EMBL; M26549; AAA53098.1; JOINED.  
CC EMBL; M26551; AAA53098.1; JOINED.  
CC EMBL; M26552; AAA53098.1; JOINED.  
CC EMBL; M26553; AAA53098.1; JOINED.  
CC EMBL; M26554; AAA53098.1; JOINED.  
CC EMBL; M26555; AAA53098.1; JOINED.  
CC EMBL; M26556; AAA53098.1; JOINED.  
CC EMBL; M26557; AAA53098.1; JOINED.  
CC EMBL; M26539; AAA53098.1; JOINED.  
CC EMBL; M26558; AAA53098.1; JOINED.  
CC EMBL; M26559; AAA53098.1; JOINED.  
CC EMBL; M26560; AAA53098.1; JOINED.  
CC EMBL; M26561; AAA53098.1; JOINED.  
CC EMBL; M26562; AAA53098.1; JOINED.  
CC EMBL; M26536; AAA53098.1; JOINED.  
CC EMBL; M26563; AAA53098.1; JOINED.  
CC EMBL; M26541; AAA53098.1; JOINED.  
CC EMBL; M26564; AAA53098.1; JOINED.  
CC EMBL; M26565; AAA53098.1; JOINED.  
CC EMBL; M26566; AAA53098.1; JOINED.  
CC EMBL; M26567; AAA53098.1; JOINED.  
CC EMBL; M26568; AAA53098.1; JOINED.  
CC EMBL; M26569; AAA53098.1; JOINED.  
CC EMBL; M26570; AAA53098.1; JOINED.  
CC EMBL; M26571; AAA53098.1; JOINED.  
CC EMBL; M26572; AAA53098.1; JOINED.  
CC EMBL; M26573; AAA53098.1; JOINED.  
CC EMBL; M26574; AAA53098.1; JOINED.  
CC EMBL; M26575; AAA53098.1; JOINED.  
CC EMBL; Y00706; CAA68698.1; JOINED.  
CC EMBL; X05561; CAA29075.1; JOINED.  
CC EMBL; M10940; AAA52006.1; JOINED.  
CC EMBL; M11315; AAA52042.1; JOINED.  
CC FIR; S16876; CGH04B.  
CC Genew; HGNC:2202; COL4A1.







```
FT isoform II).
FT /FTID=VSP_001159.
SQ SEQUENCE 1763 AA, 168526 MW, 304F528BC06AAE0D CRC64;

Query Match 76.2%; Score 115; DB 1; Length 1763;
Best Local Similarity 80.0%; Pred. No. 6.6e-09;
Matches 20; Conservative 1; Mismatches 0; Gaps 0;

QY 3 RFTTMTPLFCNVNDVCFASRNDYS 27
DB 1584 RFSTMTPLFCNVNDVCFASRNDKS 1608

RESULT 8
CA24 CAEEL STANDARD; PRT; 1758 AA.
ID CA24 CAEEL STANDARD; PRT; 1758 AA.
AC P17140; Q19098; Q19099;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (Lethal protein 2).
GN LST-2 OR CLB-1 OR F01G12.5.
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A., AND FUNCTION.
RX STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Wu X., Le T.T.;
RT "Submitted (APR-1996) to the EMBL/GenBank/DBJ databases."
RL [4]
RN [4]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMBO J. 13:3278-3285(1994).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I; Synonyms=a;
CC IsoId=P17140-1; Sequence=Displayed;
CC Name=II; Synonyms=b;
CC IsoId=P17140-2; Sequence=VSP_001160;
CC -!- DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
```

```
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 2 QRTTMTPLFCNVNDVNCNPNASNDYS 27
Db 1581 QRFSTMTPLFCNVNDVNCNPNASNDYS 1606
RESULT 9
CAL4_CABEL STANDARD; PRT; 1758 AA.
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CUB-2 OR KOH4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Cooper J., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latselle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sornhamer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RN genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.
CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Eye, and Kidney;  
RX MEDLINE=94171779; PubMed=9125972;  
RA Ohashi T., Sugimoto M., Mattei M.-G., Nimomiya Y.,  
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA  
RT isolation and assignment of the gene to chromosome Xq22, which is the  
RT same locus for COL4A5.";  
RL J. Biol. Chem. 269:7520-7526(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=94230418; PubMed=9175748;  
RA Zhou J., Ding X., Zhao Z., Readers S.T.;  
RT "Complete primary structure of the sixth chain of human basement  
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)  
RT and comparison with five other type IV collagen chains.";  
RL J. Biol. Chem. 269:13193-13199(1994).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND  
RP LYS-1110  
RX MEDLINE=96299642; PubMed=8661006;  
RA Zhang X., Zhou J., Readers S.T., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Bird C., Grahm D., Lawlor S., Wilson S.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=93361972; PubMed=9356449;  
RA Zhou J., Mochizuki T., Smets H., Antignac C., Laurila P.,  
RT de Paape A., Tryggvason K., Readers S.T.;  
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RT inherited smooth muscle tumors.";  
RL Science 261:1167-1169(1993).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequence=VSP\_001174;  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D21337; BAA04809.1; -  
CC EMBL; U04845; AAA19569.2; -  
CC EMBL; U47004; AAB19038.1; -  
CC EMBL; U46959; AAB19038.1; JOINED.  
CC EMBL; U46961; AAB19038.1; JOINED.  
CC EMBL; U46962; AAB19038.1; JOINED.  
CC EMBL; U46963; AAB19038.1; JOINED.  
CC EMBL; U46964; AAB19038.1; JOINED.  
CC EMBL; U46965; AAB19038.1; JOINED.  
CC EMBL; U46966; AAB19038.1; JOINED.  
CC EMBL; U46967; AAB19038.1; JOINED.  
CC EMBL; U46968; AAB19038.1; JOINED.  
CC EMBL; U46969; AAB19038.1; JOINED.  
CC EMBL; U46970; AAB19038.1; JOINED.  
CC EMBL; U46971; AAB19038.1; JOINED.  
CC EMBL; U46972; AAB19038.1; JOINED.  
CC EMBL; U46973; AAB19038.1; JOINED.  
CC EMBL; U46974; AAB19038.1; JOINED.  
CC EMBL; U46975; AAB19038.1; JOINED.  
CC EMBL; U46976; AAB19038.1; JOINED.  
CC EMBL; U46977; AAB19038.1; JOINED.  
CC EMBL; U46978; AAB19038.1; JOINED.  
CC EMBL; U46979; AAB19038.1; JOINED.  
CC EMBL; U46980; AAB19038.1; JOINED.  
CC EMBL; U46981; AAB19038.1; JOINED.  
CC EMBL; U46982; AAB19038.1; JOINED.  
CC EMBL; U46983; AAB19038.1; JOINED.  
CC EMBL; U46984; AAB19038.1; JOINED.  
CC EMBL; U46985; AAB19038.1; JOINED.  
CC EMBL; U46986; AAB19038.1; JOINED.  
CC EMBL; U46987; AAB19038.1; JOINED.  
CC EMBL; U46988; AAB19038.1; JOINED.  
CC EMBL; U46989; AAB19038.1; JOINED.  
CC EMBL; U46990; AAB19038.1; JOINED.  
CC EMBL; U46991; AAB19038.1; JOINED.  
CC EMBL; U46992; AAB19038.1; JOINED.  
CC EMBL; U46993; AAB19038.1; JOINED.  
CC EMBL; U46994; AAB19038.1; JOINED.  
CC EMBL; U46995; AAB19038.1; JOINED.  
CC EMBL; U46996; AAB19038.1; JOINED.  
CC EMBL; U46997; AAB19038.1; JOINED.  
CC EMBL; U46998; AAB19038.1; JOINED.  
CC EMBL; U46999; AAB19038.1; JOINED.  
CC EMBL; U47000; AAB19038.1; JOINED.  
CC EMBL; U47001; AAB19038.1; JOINED.  
CC EMBL; U47002; AAB19038.1; JOINED.  
CC EMBL; U47003; AAB19038.1; JOINED.  
CC EMBL; U47004; AAB19039.1; -  
CC EMBL; U46960; AAB19039.1; JOINED.  
CC EMBL; U46961; AAB19039.1; JOINED.  
CC EMBL; U46962; AAB19039.1; JOINED.  
CC EMBL; U46963; AAB19039.1; JOINED.  
CC EMBL; U46964; AAB19039.1; JOINED.  
CC EMBL; U46965; AAB19039.1; JOINED.  
CC EMBL; U46966; AAB19039.1; JOINED.  
CC EMBL; U46967; AAB19039.1; JOINED.  
CC EMBL; U46968; AAB19039.1; JOINED.  
CC EMBL; U46969; AAB19039.1; JOINED.  
CC EMBL; U46970; AAB19039.1; JOINED.  
CC EMBL; U46971; AAB19039.1; JOINED.  
CC EMBL; U46972; AAB19039.1; JOINED.  
CC EMBL; U46973; AAB19039.1; JOINED.  
CC EMBL; U46974; AAB19039.1; JOINED.  
CC EMBL; U46975; AAB19039.1; JOINED.  
CC EMBL; U46976; AAB19039.1; JOINED.  
CC EMBL; U46977; AAB19039.1; JOINED.  
CC EMBL; U46978; AAB19039.1; JOINED.  
CC EMBL; U46979; AAB19039.1; JOINED.  
CC EMBL; U46980; AAB19039.1; JOINED.  
CC EMBL; U46981; AAB19039.1; JOINED.  
CC EMBL; U46982; AAB19039.1; JOINED.  
CC EMBL; U46983; AAB19039.1; JOINED.





carboxyterminal, non-collagenous aggregation and cross-linking domain  
 of basement-membrane type IV collagen."  
 Eur. J. Biochem. 176:617-624 (1988).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire',  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
 CC alpha 6(IV), each of which can form a triple helix structure  
 CC with 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the  
 CC G-X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X05562; CAA29076.1; -  
 CC EMBL; X05610; CAA29098.1; -  
 CC EMBL; J02760; AAA58422.1; -  
 CC EMBL; M36963; AAA53099.1; -  
 CC EMBL; X12784; CAA31275.1; -  
 CC EMBL; J04217; AAA53097.1; -  
 CC PIR; A32024; CGHU2B.  
 CC Genew; HGNC:2203; COL4A2.  
 CC MIM; 120090; -  
 CC GO; GO:0005587; C:collagen type IV; TAS.  
 CC GO; GO:0005201; F:extracellular matrix structural constituent; TAS.  
 CC GO; GO:0030198; F:extracellular matrix organization and bioge. .; NAS.  
 CC InterPro; IPR008161; Clg helix.  
 CC InterPro; IPR008160; Collagen.  
 CC InterPro; IPR001442; Procollagn4\_C.  
 CC Pfam; PF01413; C4; 2.  
 CC Pfam; PF01391; Collagen; 24.  
 CC ProDom; PD000007; Clg\_helix; 7.  
 CC ProDom; PD003923; ProcollagnC4; 1.  
 CC SMART; SM00111; C4; 2.  
 CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25  
 FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 184 1494 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).  
 FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).  
 FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).  
 FT DISULFID 1549 1555 BY SIMILARITY.  
 FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).  
 FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).  
 FT DISULFID 1658 1665 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).  
 FT CONFLICT 471 471 R -> P (IN REF. 2).  
 FT CONFLICT 683 683 A -> I (IN REF. 2).  
 FT CONFLICT 1575 1575 M -> I (IN REF. 5).  
 FT CONFLICT 1663 1663 G -> H (IN REF. 9).  
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
 FT SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
 Query Match 67.5%; Score 102; DB 1; Length 1712;  
 Best Local Similarity 72.0%; Pred. No. 5.2e-07;

Matches 18; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 RETTPEFLFCNVNDVCNEASNDYS 27  
 DB 1540 RFTSTPEFLFCNVNDVCNEASNDYS 1564  
 ||:|||||:|||||:|||||:|||||:  
 RESULT 13  
 CA44\_RABIT  
 ID CA44\_RABIT STANDARD; PRT; 623 AA.  
 AC P55787;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Collagen alpha 4(IV) chain (Fragment).  
 DE COL4A4.  
 GN Oryctolagus cuniculus (Rabbit).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Corneal endothelium;  
 RX MEDLINE=93054733; PubMed=1429714;  
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.,  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 RT alpha 4 chain of basement membrane collagen type IV and assignment of  
 RT the gene to the distal long arm of human chromosome 2.",  
 RL J. Biol. Chem. 267:23753-23758 (1992).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire',  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
 CC alpha 6(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L01477; -; NOT\_ANNOTATED\_CDS.  
 CC PIR; A45137; A45137.  
 CC InterPro; IPR008160; Collagen.  
 CC InterPro; IPR001442; Procollagn4\_C.  
 CC Pfam; PF01413; C4; 2.  
 CC Pfam; PF01391; Collagen; 5.  
 CC ProDom; PD003923; ProcollagnC4; 1.  
 CC SMART; SM00111; C4; 2.  
 CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 392 TRIPLE-HELICAL REGION.  
 FT DOMAIN 393 623 NONHELICAL REGION (NC1).  
 FT DISULFID 413 502 OR 499 (BY SIMILARITY).  
 FT DISULFID 446 499 OR 502 (BY SIMILARITY).





```
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR001442; Procollagen4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 21.
DR ProDom: PD000007; Clg_helix; 3.
DR ProDom: PD003923; ProcollagenC4; 1.
DR SMART: SW00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1690 COLLAGEN ALPHA 4 (IV) CHAIN.
FT DOMAIN 35 64 7S DOMAIN.
FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).
FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1206 1207 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
FT DISULFID 1525 1531 BY SIMILARITY.
FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
FT DISULFID 1634 1641 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 Missing (in AS).
FT VARIANT 441 446 /FTid=VAR_008148.
FT VARIANT 545 545 G -> A (in GDSNP:1800516).
FT VARIANT 570 570 /FTid=VAR_008149.
FT VARIANT 570 570 E -> Q.
FT VARIANT 897 897 /FTid=VAR_008150.
FT VARIANT 897 897 G -> E (in FBH).
FT VARIANT 931 931 /FTid=VAR_001912.
FT VARIANT 931 931 A -> T.
FT VARIANT 1004 1004 /FTid=VAR_008151.
FT VARIANT 1004 1004 L -> P (in dbSNP:1800517).
FT VARIANT 1030 1030 /FTid=VAR_008152.
FT VARIANT 1030 1030 G -> V (in AS).
FT VARIANT 1201 1201 /FTid=VAR_008153.
FT VARIANT 1201 1201 G -> S (in AS).
FT VARIANT 1201 1201 /FTid=VAR_001913.
FT VARIANT 1402 1402 P -> S.
FT VARIANT 1572 1572 /FTid=VAR_008154.
FT VARIANT 1572 1572 P -> L (in AS).
FT CONFLICT 1659 1660 /FTid=VAR_008155.
FT CONFLICT 1659 1660 LQ -> FE (IN REF. 3).
FT SEQUENCE 1690 AA; 164095 MW; E1E72F293A72BARE CRC64;
Query Match 56.3%; Score 88; DB 1; Length 1690;
Best Local Similarity 54.2%; Pred. No. 5.8e-05;
Matches 13; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 4 FTTPMPLFCNVNDVCFNFSNDYS 27
Db 1517 FSTLPAYCNHQVCHYQNRDS 1540
RESULT 15
CA14_DROME STANDARD; PRT; 1775 AA.
AC P08130;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
```

Collagen alpha 1(IV) chain precursor.  
CG25C OR DCG1.  
Drosophila melanogaster (Fruit fly).  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Eohydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=89054012; PubMed=3142875;  
Blumberg B., Mackrell A.J., Fessler J.H.;  
"Drosophila basement membrane procollagen alpha 1(IV). II. Complete  
cDNA sequence, genomic structure, and general implications for  
supramolecular assemblies.";  
J. Biol. Chem. 263:18328-18337(1988).  
[2]  
SEQUENCE FROM N.A.  
Blumberg B.;  
Thesis (1987), University of California / Los Angeles, U.S.A.  
[3]  
SEQUENCE FROM N.A.  
Mackrell A.J.;  
Thesis (1992), University of California / Los Angeles, U.S.A.  
[4]  
SEQUENCE OF 1065-1775 FROM N.A.  
MEDLINE=87194801; PubMed=3106346;  
Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,  
Natzle J.E., Fessler J.H.;  
"Basement membrane procollagen IV and its specialized carboxyl domain  
are conserved in Drosophila, mouse, and human.";  
J. Biol. Chem. 262:5947-5950(1987).  
[5]  
SEQUENCE OF 1355-1775 FROM N.A.  
MEDLINE=87246644; PubMed=3109906;  
Cecchini J.P., Knibiehler B., Mirre C., Le Parco Y.;  
"Evidence for a type-IV-related collagen in Drosophila melanogaster.  
Evolutionary constancy of the carboxyl-terminal noncollagenous  
domain.";  
Eur. J. Biochem. 165:587-593(1987).  
[6]  
SEQUENCE OF 762-1230 FROM N.A.  
MEDLINE=82197577; PubMed=6210912;  
Monson J.M., Natzle J., Friedman J., McCarthy B.J.;  
"Expression and novel structure of a collagen gene in Drosophila.";  
Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).  
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.  
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
CC Type IV collagen forms a mesh-like network linked through  
intermolecular interactions between 7S domains and between NC1  
domains.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the G-  
X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
CC -!- PM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
EMBL; M23704; AAA28404.1; -;  
DR EMBL; M96575; AAB59184.1; -;  
DR EMBL; J02727; AAA28423.1; -;

DR EMBL; M28334; AAA28422.1; -.  
DR EMBL; V00200; CAA23486.2; -.  
DR PIR; A31893; A31893  
DR FlyBase; FBgn0000299; Cg25C.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 25.  
DR ProDom; PDOC00007; C1g\_helix; 9.  
DR ProDom; PDOC03923; ProcollagenC4; 1.  
DR SMART; SMO0111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 23  
FT PROPEP 24 ?  
FT CHAIN ? 1775 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT DOMAIN ? 1544 COLLAGEN ALPHA 1(IIV) CHAIN.  
FT DOMAIN 1545 1775 TRIPLE-HELICAL REGION.  
FT DISULFID 1569 1655 NONHELICAL REGION (NC1).  
FT DISULFID 1599 1652 OR 1652 (BY SIMILARITY).  
FT DISULFID 1611 1617 OR 1655 (BY SIMILARITY).  
FT DISULFID 1674 1770 BY SIMILARITY.  
FT DISULFID 1708 1767 OR 1767 (BY SIMILARITY).  
FT DISULFID 1720 1727 OR 1770 (BY SIMILARITY).  
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CONFLICT 948 948 L -> S (IN REF. 6).  
FT CONFLICT 997 997 S -> T (IN REF. 6).  
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).  
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).  
FT CONFLICT 1373 1373 T -> I (IN REF. 5).  
FT CONFLICT 1496 1496 L -> R (IN REF. 5).  
FT CONFLICT 1507 1511 ETGNV -> RAGOR (IN REF. 5).  
FT CONFLICT 1529 1529 E -> K (IN REF. 5).  
FT CONFLICT 1733 1733 M -> I (IN REF. 5).  
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AB23149525CD CRC64;

Query Match 58.3%; Score 88; DB 1; Length 1775;  
Best Local Similarity 65.2%; Pred. No. 6.1e-05;  
Matches 15; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 RPTMPFLFCNVNDYCNESRND 25  
Db 1602 RFSTLPVLSCGQNNVCNYSRND 1624

Search completed: April 5, 2004, 06:59:39  
Job time : 4.39952 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 Seconds  
(without alignments)  
525.440 Million cell updates/sec

Title: US-10-032-221b-39

Perfect score: 151

Sequence: 1 KQRTTTPFLFCNVNDVCNCFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	96.7	212	6 Q28512	Q28512 macaca mula
2	146	96.7	245	4 Q9NYC4	Q9NYC4 homo sapien
3	145	96.0	203	6 Q29032	Q29032 sus scrofa
4	145	96.0	203	6 Q28682	Q28682 oryctolagus
5	145	96.0	212	6 Q28567	Q28567 ovis aries
6	140	92.7	161	11 Q61430	Q61430 mus musculus
7	140	92.7	210	6 Q28273	Q28273 canis famil
8	140	92.7	246	11 Q61435	Q61435 mus musculus
9	140	92.7	1669	11 Q9QZS0	Q9QZS0 mus musculus
10	136	90.1	230	11 Q63122	Q63122 rattus norv
11	130	86.1	179	11 P70165	P70165 mus musculus
12	130	86.1	253	11 Q61436	Q61436 mus musculus
13	130	86.1	585	11 Q80V57	Q80V57 mus musculus
14	130	86.1	799	11 Q8BNS7	Q8BNS7 mus musculus
15	130	86.1	886	4 Q9NUB7	Q9NUB7 homo sapien
16	130	86.1	1684	6 Q8HYC1	Q8HYC1 canis famil

17	130	86.1	1688	6 Q85622	Q85622 canis famil
18	130	86.1	1891	11 Q8ESQ2	Q8ESQ2 mus musculus
19	129	85.4	225	6 Q28271	Q28271 canis famil
20	129	85.4	226	11 Q99LQ8	Q99LQ8 mus musculus
21	129	85.4	229	4 Q8NF88	Q8NF88 homo sapien
22	129	85.4	229	4 Q9NYC5	Q9NYC5 homo sapien
23	129	85.4	979	13 Q919K3	Q919K3 gallus gall
24	129	85.4	1075	4 Q86X41	Q86X41 homo sapien
25	129	85.4	1821	4 Q9H4R9	Q9H4R9 homo sapien
26	120	79.5	1747	5 Q28640	Q28640 strongyloce
27	120	79.5	1752	5 Q07265	Q07265 strongyloce
28	108	71.5	1802	5 Q17163	Q17163 brugia mala
29	105	69.5	205	6 Q28274	Q28274 canis famil
30	105	69.5	546	11 Q99K97	Q99K97 mus musculus
31	105	69.5	1600	4 Q9UEH6	Q9UEH6 homo sapien
32	103	69.5	1691	11 Q8ESQ1	Q8ESQ1 mus musculus
33	102	67.5	202	6 Q28272	Q28272 canis famil
34	102	67.5	358	11 Q91VI3	Q91VI3 mus musculus
35	102	67.5	673	4 Q14052	Q14052 homo sapien
36	99	65.6	1723	5 Q9GQB1	Q9GQB1 hydra atten
37	89	58.9	1761	5 Q18407	Q18407 drosophila
38	89	58.9	1940	5 Q9VMV5	Q9VMV5 drosophila
39	88	58.3	312	11 Q64457	Q64457 mus musculus
40	88	58.3	1682	11 Q9QZR9	Q9QZR9 mus musculus
41	88	58.3	1779	5 Q9VMV4	Q9VMV4 drosophila
42	87	57.6	208	6 Q29468	Q29468 canis famil
43	87	57.6	1024	5 Q8T7S4	Q8T7S4 anopheles g
44	84	55.6	713	5 Q9GV24	Q9GV24 sarcophaga
45	52	34.4	177	2 Q9A1Z4	Q9A1Z4 carsonella

#### ALIGNMENTS

#### RESULT 1

Q28512 ID Q28512 PRELIMINARY; PRT; 212 AA.  
AC Q28512;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,  
Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
RT mammals";  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L47280; AAA91861.1; -;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollag\_n4\_C.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR PROSITE; PS003923; Procollag\_n4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1  
FT NON\_TER 212  
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357564 CRC64;

Query Match 96.7%; Score 146; DB 6; Length 212;  
Best Local Similarity 100.0%; Pred. No. 3.2e-15;

```
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 2
Q9NYC4 PRELIMINARY; PRT; 245 AA.
ID Q9NYC4
AC Q9NYC4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Tumorstatin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA Erickson M.D., Hoffer H., Xiao Y., Stillman I.E., Kalluri R.,
RT "Distinct anti-tumor properties of a type IV collagen domain derived
RL J. Biol. Chem. 0:0-0(2000).
DR ENBL; AF258351; AAF72632.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR ProSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 96.7%; Score 146; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 70 QRTTTPFLFCNVNDVCFASRNDYS 95

RESULT 3
Q29032 PRELIMINARY; PRT; 203 AA.
ID Q29032
AC Q29032;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RL mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR ProSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
SQ SEQUENCE 203 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 96.0%; Score 145; DB 6; Length 203;
Best Local Similarity 96.2%; Pred. No. 4.5e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 4
Q28682 PRELIMINARY; PRT; 203 AA.
ID Q28682
AC Q28682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RL mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR ProSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 96.0%; Score 145; DB 6; Length 203;
Best Local Similarity 96.2%; Pred. No. 4.5e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 5
Q28567 PRELIMINARY; PRT; 212 AA.
ID Q28567
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RL mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

```
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR ProSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER
FT NON_TER
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 96.0%; Score 145; DB 6; Length 203;
Best Local Similarity 96.2%; Pred. No. 4.5e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 4
Q28682 PRELIMINARY; PRT; 203 AA.
ID Q28682
AC Q28682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RL mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4.C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR ProSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 96.0%; Score 145; DB 6; Length 203;
Best Local Similarity 96.2%; Pred. No. 4.5e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

RESULT 5
Q28567 PRELIMINARY; PRT; 212 AA.
ID Q28567
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RL mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR ENBL; L47284; AAA91882.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
```

```

GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR DR InterPro; IPR001442; Procollagn4 C.
DR DR Pfam; PF01413; C4; 2.
DR DR InterPro; IPR000504; RNA_rec_mot.
DR DR SMART; SMO0111; C4; 2.
DR DR PROSITE; PS00030; RRM_RNP_1; 1.
DR DR COLLAGEN.
KW Collagen.
FT NON_TER 1 212
FT NON_TER 212 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BD8C CRC64;

Query Match 96.0%; Score 145; DB 6; Length 212;
Best Local Similarity 96.2%; Pred. No. 4.7e-15;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||
DB 37 QRTTTPFLFCNVNDVNCVFASRNDYS 62

RESULT 6
Q61430 Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the Ncl domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR DR InterPro; IPR001442; Procollagn4 C.
DR DR Pfam; PF01413; C4; 2.
DR DR SMART; SMO0111; C4; 2.
DR DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1 161
FT NON_TER 161 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

```

```

Query Match 92.7%; Score 140; DB 11; Length 161;
Best Local Similarity 92.3%; Pred. No. 2.2e-14;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||
DB 4 QRTTTPFLFCNVNDVNCVFASRNDYS 29

RESULT 7
Q28273 Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng X., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR DR InterPro; IPR001442; Procollagn4 C.
DR DR Pfam; PF01413; C4; 2.
DR DR InterPro; IPR000504; RNA_rec_mot.
DR DR SMART; SMO0111; C4; 1.
DR DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1 210
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA82363D CRC64;

Query Match 92.7%; Score 140; DB 6; Length 210;
Best Local Similarity 92.3%; Pred. No. 2.9e-14;
Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCVFASRNDYS 27
|||||:|||||:|||||:|||||
DB 47 QRTTTPFLFCNVNDVNCVFASRNDYS 72

RESULT 8
Q61435 Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```

RT laminae: Sequence, distribution, association with laminins, and  
 RT developmental switches.";  
 RL J. Cell Biol. 127:879-891(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Balb/c;  
 RA Miner J.H.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z35166; CAA84529.1; -;  
 DR PIR; I48302; I48302.  
 DR MGD; MGI:104688; Col4a3.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR001442; Procollagn4.C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 FT NON TER 1  
 SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;  
 Query Match 92.7%; Score 140; DB 11; Length 246;  
 Best Local Similarity 92.3%; Pred. No. 3.4e-14;  
 Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QRTTTPFLFCNVNDVNCNFSARNDS 27  
 DB 71 QRTTTPFLFCNVNDVNCNFSARNDS 96  
 RESULT 9  
 Q9QZS0  
 ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.  
 AC Q9QZS0;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Alpha 3 collagen IV.  
 GN COL4A3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,  
 RA Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;  
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a  
 RT mouse model of alport syndrome.";  
 RL Genomics 61:113-124(1999).  
 DR EMBL; AF169387; AAD50449.1; -;  
 DR PIR; I48302; I48302.  
 DR MGD; MGI:104688; Col4a3.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagn4.C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000007; C1g\_helix; 6.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Collagen.  
 SQ SEQUENCE 1669 AA; 161769 MW; 30976B59739A47B2 CRC64;  
 Query Match 92.7%; Score 140; DB 11; Length 1669;  
 Best Local Similarity 92.3%; Pred. No. 2.3e-13;  
 Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVNCNFSARNDS 27  
 DB 1494 QRTTTPFLFCNVNDVNCNFSARNDS 1519  
 RESULT 10  
 Q63122  
 ID Q63122 PRELIMINARY; PRT; 230 AA.  
 AC Q63122;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).  
 GN COL4A3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;  
 RA MEDLINE=98210005; PubMed=950634;  
 RA Ryan J.J., Katbama I., Mason P.J., Pusey C.D., Turner A.N.;  
 RT "Sequence analysis of the 'Goodpasture antigen' of mammals.";  
 RL Nephrol. Dial. Transplant. 13:602-607(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;  
 RA Turner N.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; I47281; AAE72238.2; -;  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001442; Procollagn4.C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Collagen.  
 FT NON TER 1  
 FT NON TER 230  
 SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;  
 Query Match 90.1%; Score 136; DB 11; Length 230;  
 Best Local Similarity 92.3%; Pred. No. 1.4e-13;  
 Matches 24; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QRTTTPFLFCNVNDVNCNFSARNDS 27  
 DB 55 QRTTTPFLFCNVNDVNCNFSARNDS 80  
 RESULT 11  
 P70165  
 ID P70165 PRELIMINARY; PRT; 179 AA.  
 AC P70165;  
 DT 01-FEB-1997 (TReMBLrel. 02, Created)  
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Collagen type IV alpha5 chain (Fragment).  
 GN COL4A5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RA Oberbauer I.;  
 RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse  
 RT via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and

```

RT alphas(IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82218; CAA57698.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR DR InterPro; IPR001442; Procollagn4_C.
DR DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMC0111; C4; 2.
FT NON TER 179 179
FT NON TER 179 179
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 86.1%; Score 130; DB 11; Length 179;
Best Local Similarity 80.8%; Pred. No. 9.7e-13;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTTPELFNCNVNDVNCNFSARNDIS 27
Db 32 RRFSTMPFMCNINNVNCFASRNDIS 57

RESULT 12
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.B., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043317; AAH43317.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01413; C4; 2.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
DR SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 86.1%; Score 130; DB 11; Length 585;
Best Local Similarity 80.8%; Pred. No. 3.2e-12;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTTPELFNCNVNDVNCNFSARNDIS 27
Db 411 RRFSTMPFMCNINNVNCFASRNDIS 436

RESULT 14
Q8BNS7 PRELIMINARY; PRT; 799 AA.
AC Q8BNS7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Procollagen (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```



```
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AK080882; BAC37980.1; -.
DR MGD; MGI:188456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 86.1%; Score 130; DB 11; Length 799;
Best Local Similarity 80.8%; Pred.No. 4.8e-12;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVNCNFSARNDS 27
   :|||:|||||:|||||:|||||
Db 625 RRFSTMPFMCNINNVCNFSARNDS 650

RESULT 15
Q9NUTB7
AC Q9NUTB7 PRELIMINARY; PRT; 886 AA.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
DE DE
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 86.1%; Score 130; DB 4; Length 886;
Best Local Similarity 80.8%; Pred.No. 4.8e-12;
Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVNCNFSARNDS 27
   :|||:|||||:|||||:|||||
Db 712 RRFSTMPFMCNINNVCNFSARNDS 737

Search completed: April 5, 2004, 07:03:58
Job time : 17.2131 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221b-39  
Perfect score: 151  
Sequence: 1 KQFTTTPFLFCNVNDVCFASNDYS 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	27	6	ADA20238
2	146	96.7	79	5	AAU75600 Human typ
3	146	96.7	79	6	ADA20264 Human typ
4	146	96.7	88	5	AAU75608 Human typ
5	146	96.7	88	5	AAU75607 Human typ
6	146	96.7	88	6	ADA20271 Human typ
7	146	96.7	88	6	ADA20272 Human typ
8	146	96.7	124	5	AAU75594 Human typ
9	146	96.7	124	6	ADA20258 Human typ
10	146	96.7	132	5	AAU75597 Human typ
11	146	96.7	132	6	ADA20261 Human typ
12	146	96.7	191	5	AAU75596 Human typ
13	146	96.7	191	6	ADA20260 Human typ
14	146	96.7	211	3	AAU95918 Human typ
15	146	96.7	211	5	ABG79208 Human GP
16	146	96.7	218	2	AAU79164 Human typ
17	146	96.7	218	2	AAU44172 Human typ
18	146	96.7	218	3	AAU56784 Human typ
19	146	96.7	218	4	AAE09484 Human typ
20	146	96.7	232	7	ADC17697 Human typ
21	146	96.7	244	5	ABG79218 Human typ
22	146	96.7	244	5	ABG79219 Human typ
23	146	96.7	244	5	ABG79217 Human typ
24	146	96.7	244	5	AAU75595 Human typ
25	146	96.7	244	6	ADA20225 Human typ

26	146	96.7	245	3	AAU75942	Human typ
27	146	96.7	245	5	AAU75589	Human typ
28	146	96.7	254	5	AAU75598	Human typ
29	146	96.7	268	3	AAU31993	Type IV C
30	146	96.7	268	3	AAU97555	Human alp
31	146	96.7	1670	7	ADA47063	Human pro
32	145	96.0	471	2	AAU79163	Partial s
33	145	96.0	471	2	AAU44171	Bovine ty
34	145	96.0	471	3	AAU56783	Bovine al
35	145	96.0	471	4	AAE09483	Bovine al
36	136	90.1	230	7	ADA47061	Rat Prote
37	131	86.8	27	6	ADA20239	T8-3 pept
38	130	86.1	229	7	ADC17699	Human typ
39	130	86.1	264	2	AAU31995	Type IV C
40	130	86.1	264	3	AAU97557	Human alp
41	130	86.1	309	3	AAU54044	Human pan
42	130	86.1	772	2	AAU23873	Human alp
43	130	86.1	772	2	AAU09643	Human typ
44	130	86.1	1685	4	ABG04839	Novel hum
45	130	86.1	1693	4	ABG15619	Novel hum

## ALIGNMENTS

## RESULT 1

ADA20238  
ID ADA20238 standard; peptide; 27 AA.

XX ADA20238;

XX 20-NOV-2003 (first entry)

DE T8 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network; C-terminal globular non-collagenous domain; NC1; type IV collagen; cell surface receptor; integrin; angiogenic activity; protein synthesis; cytostatic; gene therapy; T8 peptide; tumstatin; human; type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-597256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 62; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the T8 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.

XX SQ Sequence 27 AA;

Query Match 100.0%; Score 151; DB 6; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-16;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFCNVNDVCFASRNDYS 27  
 |||||  
 Db 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

RESULT 2

AAU75600  
 ID AAU75600 standard; protein; 79 AA.

XX AC AAU75600;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tum-5.

XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutelin; mutant.

XX OS Homo sapiens.

XX FN WO200151523-A2.

XX PD 19-JUL-2001.

XX XX 08-JAN-2001; 2001WO-US000565.

XX PF 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX XX Kalluri R;

XX XX WPI; 2002-188037/24.

XX XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.

XX PS Example 40; Page; 205pp; English.

XX XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues  
 CC 54-132 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human tumstatin sequence  
 CC given in figure 18A (see AAU75589)

XX SQ Sequence 79 AA;

Query Match 96.7%; Score 146; DB 5; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-15;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFCNVNDVCFASRNDYS 27  
 |||||  
 Db 17 QRFTTTPFLFCNVNDVCFASRNDYS 42

RESULT 3

ADA20264

XX ADA20264 standard; protein; 79 AA.

XX AC ADA20264;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-5 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX XX 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.  
DR N-PSDB; ADA20224.  
XX  
XX  
PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
PS Claim 94; SEQ ID NO 26; 240pp; English.  
XX  
CC This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumor growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
CC the invention which was derived from the amino acid sequence of the alpha  
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
CC not appear in the specification but was created by the indexer from  
CC information given in the specification.  
XX  
SQ Sequence 79 AA;  
XX  
Query Match 96.7%; Score 146; DB 6; Length 79;  
Best Local Similarity 100.0%; Pred. No. 6.9e-15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTMPFLFCNVNDVCFNSRNDYS 27  
DB 16 QRFTMPFLFCNVNDVCFNSRNDYS 41  
XX  
RESULT 4  
AAU75608  
ID AAU75608 standard; protein; 88 AA.  
XX  
AC AAU75608;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.  
XX  
KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 82  
ET /note= "Wild type Cys substituted with Ala"  
XX  
FN WO200151523-A2.  
XX  
PD 19-JUL-2001.  
XX  
XX 08-JAN-2001; 2001WO-US0000565.  
XX  
FR 07-JAN-2000; 2000US-00479118.  
FR 04-APR-2000; 2000US-00543371.  
FR 21-JUL-2000; 2000US-00625191.  
XX  
XX

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
Kalluri R;  
WPI; 2002-188037/24.  
A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
treating disorders involving angiogenesis.  
Claim 41; Page 153; 205pp; English.  
The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
domain, having one or more of the characteristics selected from: (a) the  
ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
proliferation of endothelial cells; and (c) the ability to cause  
apoptosis of endothelial cells. Also described are the following: (1) use  
of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
analogue or allelic variant in the preparation of a medicament for  
treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
where the angiogenesis is mediated by one or more endothelial cell  
integrins or one or more endothelial cell integrin subunits; or (b) by  
promoting or inducing endothelial cell apoptosis in a tissue, where the  
endothelial cell apoptosis is mediated by one or more endothelial cell  
integrins or one or more endothelial cell integrin subunits; (2) use of  
alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
preparation of a medicament for inhibiting angiogenesis or cell  
proliferation; (3) use of an inhibitor, such as an antibody, antibody  
fragment or peptide of receptor-mediated angiogenesis in the preparation  
of a medicament for treating a proliferative disease in a vertebrate,  
where the disease is characterised by angiogenesis that is mediated by  
receptors to Arresten, Canstatin or Tumstatin and where the receptors  
inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
the presence of a medicament for promoting angiogenesis in a tissue; and  
(5) use of integrins in the preparation of a medicament for promoting or  
inducing angiogenesis or cell proliferation in a tissue. The fragments  
of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
or allelic variants are useful in the preparation of a medicament for  
treating a disorder involving inhibiting angiogenesis in a tissue, where  
the angiogenesis is mediated by one or more endothelial cell integrins or  
one or more endothelial cell integrin subunits; or by promoting or  
inducing endothelial cell apoptosis in a tissue, where the endothelial  
cell apoptosis is mediated by one or more endothelial cell integrins or  
one or more endothelial cell integrin subunits. The medicament is useful  
in inhibiting tumour growth and for the regression of an established  
tumour. The present sequence represents the amino acid sequence of human  
type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which  
consists of residues 5-126 of Tumstatin  
XX  
SQ Sequence 88 AA;  
XX  
Query Match 96.7%; Score 146; DB 5; Length 88;  
Best Local Similarity 100.0%; Pred. No. 7.8e-15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTMPFLFCNVNDVCFNSRNDYS 27  
DB 26 QRFTMPFLFCNVNDVCFNSRNDYS 51  
XX  
RESULT 5  
AAU75607  
ID AAU75607 standard; protein; 88 AA.  
XX  
AC AAU75607;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.  
XX  
KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW

KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
OS Homo sapiens.  
XX WO200151523-A2.  
XX 19-JUL-2001.  
XX 08-JAN-2001; 2001WO-US000565.  
XX 07-JAN-2000; 2000US-00479118.  
PR 04-APR-2000; 2000US-00543371.  
PR 21-JUL-2000; 2000US-00625191.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX Claim 32; Page 152; 205pp; English.  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterized by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin and where the receptors  
CC are more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists  
CC of residues 45-132 of Tumstatin  
XX Sequence 88 AA;  
Query Match 96.7%; Score 146; DB 5; Length 88;  
Best Local Similarity 100.0%; Pred. NO. 7.8e-15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFCNVNDVNCNFASRNDYS 27  
|||||

Db 26 QRFTHMPFLFCNVNDVNCNFASRNDYS 51  
RESULT 6  
ADA20271  
ID ADA20271 standard; protein; 88 AA.  
XX ADA20271;  
AC ADA20271;  
XX 20-NOV-2003 (first entry)  
XX Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.  
DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human;  
KW tumstatin 45-132.  
XX Homo sapiens.  
OS WO2003059257-A2.  
XX 24-JUL-2003.  
XX 20-DEC-2002; 2002WO-US040938.  
XX 21-DEC-2001; 2001US-00032221.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2003-587256/55.  
XX N-PSDB; ADA20224.  
XX New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX Claim 94; SEQ ID NO 33; 240pp; English.  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
CC ID33) does not appear in the specification but was created by the indexer  
CC from information given in the specification.  
XX Sequence 88 AA;  
Query Match 96.7%; Score 146; DB 6; Length 88;  
Best Local Similarity 100.0%; Pred. NO. 7.8e-15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFCNVNDVNCNFASRNDYS 27  
|||||

RESULT 7  
ADA20272  
ID ADA20272 standard; protein; 88 AA.  
XX AC  
XX AC  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human;  
XX KW tumstatin 5-125-C-A; mutant; mutein.  
XX OS Synthetic.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 81 /note= "wild-type Cys substituted by Ala at position 125  
XX FT of full-length tumstatin"  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 34; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
XX CC with anti-angiogenic properties. The invention also relates to the DNA  
XX CC sequences which encode the novel proteins. A wide variety of diseases are  
XX CC the result of undesirable angiogenesis. The formation of new capillaries  
XX CC from pre-existing vessels is essential for tumour growth and metastasis.  
XX CC Basement membrane organisation is dependent on the assembly of a type IV  
XX CC collagen network which may occur through the C-terminal globular non-  
XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX CC forms are ubiquitously exhibited in human basement membranes. In the  
XX CC present invention, cell surface receptors (in particular integrins) which  
XX CC specifically bind anti-angiogenic proteins and peptides (in particular  
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX CC sequences of the invention may be useful in gene therapy. The present  
XX CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of  
XX CC the "tumstatin" protein of the invention which was derived from the amino  
XX CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
XX CC sequence (Seq ID33) does not appear in the specification but was created  
XX CC by the indexer from information given in the specification.  
XX SQ Sequence 88 AA;  
Query Match 96.7%; Score 146; DB 6; Length 88;  
Best Local Similarity 100.0%; Pred. No. 7.8e-15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 QRETTMPFLFCNVNDVCFASRDYS 27  
|||||  
25 QRETTMPFLFCNVNDVCFASRDYS 50  
RESULT 8  
AAU75594  
ID AAU75594 standard; protein; 124 AA.  
XX AC  
XX AC  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.  
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX OS Homo sapiens.  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PR 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2002-198037/24.  
XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX PT treating disorders involving angiogenesis.  
XX PS Example 33; Page; 205pp; English.  
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX CC domain, having one or more of the characteristics selected from: (a) the  
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX CC proliferation of endothelial cells; and (c) the ability to cause  
XX CC apoptosis of endothelial cells. Also described are the following: (1) use  
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX CC analogue or allelic variant in the preparation of a medicament for  
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX CC where the angiogenesis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by  
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of  
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,  
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX CC preparation of a medicament for inhibiting angiogenesis or cell  
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX CC of a medicament for treating a proliferative disease in a vertebrate,  
XX CC where the disease is characterised by angiogenesis that is mediated by  
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and  
XX CC (5) use of integrins in the preparation of a medicament for promoting or  
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
XX CC or allelic variants are useful in the preparation of a medicament for  
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting of  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of  
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in  
CC the specification but is derived from the wild type human Tumstatin  
CC sequence given in figure 18A (see AAU75589)

XX  
SQ Sequence 124 AA;

Query Match 96.7%; Score 146; DB 5; Length 124;

Best Local Similarity 100.0%; Pred. No. 1.1e-14;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNVCNPFASRNDYS 27

|||||  
Db 69 QRFTMPFLFCNVNVCNPFASRNDYS 94

RESULT 9

ADA20258

ID ADA20258 standard; protein; 124 AA.

XX

AC ADA20258;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

XX

KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX

KW metastasis; basement membrane organisation; type IV collagen network;

XX

KW C-terminal globular non-collagenous domain; NCI; type IV collagen;

XX

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX

KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

XX

OS Homo sapiens.

XX

PN WO2003059257-A2.

XX

PD 24-JUL-2003.

XX

PF 20-DEC-2002; 2002WO-US040938.

XX

PR 21-DEC-2001; 2001US-00032221.

XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX

PI Kalluri R;

XX

WPI; 2003-587256/55.

XX

DR N-PSDB; ADA20224.

XX

PT New peptide, useful for preparing a composition for inhibiting tumor

XX

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX

PS Claim 94; SEQ ID NO 20; 240pp; English.

XX

XX

CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
CC ID20) does not appear in the specification but was created by the indexer  
CC from information given in the specification.

XX  
SQ Sequence 124 AA;

Query Match 96.7%; Score 146; DB 6; Length 124;

Best Local Similarity 100.0%; Pred. No. 1.1e-14;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNVCNPFASRNDYS 27

|||||  
Db 69 QRFTMPFLFCNVNVCNPFASRNDYS 94

RESULT 10

AAU75597

ID AAU75597 standard; protein; 132 AA.

XX

AC AAU75597;

XX

DT 08-MAY-2002 (first entry)

XX

DE Human type IV collagen alpha 3 chain mutant, Tum-2.

XX

KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

XX

KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

XX

KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;

XX

KW Tumstatin; angiogenesis; tumour; mutant.

XX

OS Homo sapiens.

XX

PN WO200151523-A2.

XX

PD 19-JUL-2001.

XX

PF 08-JAN-2001; 2001WO-US000565.

XX

PR 07-JAN-2000; 2000US-00479118.

XX

PR 04-APR-2000; 2000US-00543371.

XX

PR 21-JUL-2000; 2000US-00625191.

XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX

PI Kalluri R;

XX

WPI; 2002-188037/24.

XX

DR

PT

PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

XX

PT treating disorders involving angiogenesis.

XX

PS Claim 31; Page 152; 205pp; English.

XX

XX

CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause

CC apoptosis of endothelial cells. Also described are the following: (1) use

CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for

CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by

CC promoting or inducing endothelial cell apoptosis in a tissue, where the

CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of

CC an antibody or peptide that specifically binds the alpha1, alpha2,



CC alpha3, alpha5, alpha6, alphav, betal or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues  
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75589)

XX  
 SQ Sequence 132 AA;

Query Match 96.7%; Score 146; DB 5; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
 Db 70 QRFTHMPFLFCNVNDVCFASRNDYS 95  
 |||||

RESULT 11

ID ADA20261  
 XX ADA20261 standard; protein; 132 AA.

AC ADA20261;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor  
 growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.

XX Sequence 132 AA;

Query Match 96.7%; Score 146; DB 6; Length 132;

Best Local Similarity 100.0%; Pred. No. 1.2e-14;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27

Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94  
 |||||

RESULT 12

AAU75596

ID AAU75596 standard; protein; 191 AA.

AC AAU75596;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3 (IV)NCI domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US0000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3 (IV)NCI domain used in detecting and  
 treating disorders involving angiogenesis.

XX Example 32; Page; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or  
 CC alpha3, alpha4, alpha5, alpha6, beta1 or beta2 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate;  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of  
 CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in  
 CC the specification but is derived from the wild type human Tumstatin  
 CC sequence given in figure 10A (see AAU75589)

XX Sequence 191 AA;

Query Match 96.7%; Score 146; DB 5; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNDVCFASNDYS 27  
 DB 17 QRFTMPFLFCNVNDVCFASNDYS 42

RESULT 13

ADA20260  
 ID ADA20260 standard; protein; 191 AA.

XX ADA20260;  
 AC ADA20260;

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-1 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cyrostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;  
 KW tumstatin N53.

XX Homo sapiens.

OS WO2003059257-A2.

PN

XX 24-JUL-2003.  
 PD 20-DEC-2002; 2002WO-US040938.  
 XX 21-DEC-2001; 2001US-00032221.  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA Kalluri R;  
 XX WPI; 2003-587256/55.  
 DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX Claim 94; SEQ ID NO 22; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-1 (tumstatin N53), an abridged form of the  
 CC "tumstatin" protein of the invention which was derived from the amino  
 CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
 CC sequence (Seq ID22) does not appear in the specification but was created  
 CC by the indexer from information given in the specification.

XX Sequence 191 AA;

Query Match 96.7%; Score 146; DB 6; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTMPFLFCNVNDVCFASNDYS 27  
 DB 16 QRFTMPFLFCNVNDVCFASNDYS 41

RESULT 14

AA95918

ID AA95918 standard; protein; 211 AA.

XX AA95918;  
 AC AA95918;

DT 20-NOV-2000 (first entry)

XX Human Goodpasture antigen Deltav.

XX Goodpasture antigen; GPDeltav; goodpasture antigen binding protein; GBBP;  
 KW human; autoimmune disease; apoptosis; cancer; tumour; therapy.

XX Homo sapiens.

OS WO200050607-A2.

XX 31-AUG-2000.

XX 24-FEB-2000; 2000WO-IB000324.

XX

PT 24-FEB-1999; 99US-0121483P.  
XX (SAUS/) SAUS J.  
XX PI Saus J;  
XX WPI; 2000-572094/53.  
DR N-PSDB; AAA50367.  
XX Novel Goodpasture antigen binding proteins useful for diagnosing and  
PT treating autoimmune disorders, tumor, and preventing cell apoptosis.  
XX Claim 36; Page 151-152; 158pp; English.  
XX The present sequence is that of human recombinant Goodpasture antigen  
CC (GP) Deltav, i.e. an alternative form of human GP resulting from splicing  
CC out of exon V. The recombinant protein, lacking the Met-1 residue, was  
CC expressed in bacterial pellets using modified vector pET15b carrying  
CC GPdeltav cDNA (see AAA50367). The invention relates to novel Goodpasture  
CC antigen binding proteins (GPPs, see AA95900-11), which bind to and  
CC phosphorylate the unique N-terminal region of human GP, and which are  
CC highly expressed in several autoimmune conditions. Claimed methods for  
CC treating an autoimmune disorder, cell apoptosis or a tumour involve  
CC modifying the expression or activity of GPP, especially using a GP-  
CC derived peptide, such as GPdeltav  
XX Sequence 211 AA;  
SQ

Query Match 96.7%; Score 146; DB 3; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.1e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
DB 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 15  
ABG79208  
ID ABG79208 standard; protein; 211 AA.  
XX AC ABG79208;  
XX DT 15-NOV-2002 (first entry)  
XX DE Human GP protein isoform GPdeltav.  
XX KW Goodpasture antigen binding protein; Goodpasture syndrome;  
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;  
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;  
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
XX pemphigoid; lichen planus; human.  
XX OS Homo sapiens.  
XX FN WO200261430-A2.  
XX PD 08-AUG-2002.  
XX PF 31-JAN-2002; 2002WO-EP001010.  
XX PR 31-JAN-2001; 2001US-0265249P.  
XX (SAUS/) SAUS J.  
XX PI Saus J;  
XX WPI; 2002-619280/66.  
DR N-PSDE; ABS64491.  
XX Identifying candidate compounds for treating autoimmune conditions, e.g.  
PT Goodpasture syndrome or lupus, comprises identifying compounds that

PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.  
XX Example 3; Page 199-200; 217pp; English.  
XX The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) which is selected from Goodpasture antigen  
CC binding protein (GPP), an alpha3 type IV collagen non-collagenous (NCI)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of  
CC conformational isomers of the second target protein (II) (selected from  
CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC NCI domain conformational isomer, which has an amino acid sequence  
CC identical to the wild type alpha3 type IV collagen NCI domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kDa, and in  
CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated  
CC type IV collagen alpha3 NCI domain. The human gene for GPP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)  
CC domain (also known as the GP antigen) or an MBP isoform  
XX Sequence 211 AA;  
SQ

Query Match 96.7%; Score 146; DB 5; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.1e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
DB 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

Search completed: April 5, 2004, 06:58:32  
Job time : 25.3196 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds  
(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-39

Perfect score: 151

Sequence: 1 KQRTTTPFLFCNVNDVCFNSRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA.\*  
1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	27	14	US-10-032-221B-39
2	146	96.7	79	14	Sequence 26, Appl
3	146	96.7	88	14	Sequence 33, Appl
4	146	96.7	88	14	Sequence 34, Appl
5	146	96.7	124	14	Sequence 20, Appl
6	146	96.7	132	14	Sequence 23, Appl
7	146	96.7	191	14	Sequence 22, Appl
8	146	96.7	211	14	Sequence 46, Appl
9	146	96.7	211	14	Sequence 46, Appl
10	146	96.7	232	14	Sequence 304, Appl
11	146	96.7	244	14	Sequence 10, Appl
12	131	86.8	27	14	Sequence 40, Appl
13	130	86.1	229	9	Sequence 306, Appl
14	130	86.1	309	9	Sequence 496, Appl
15	129	85.4	229	14	Sequence 302, Appl

16	129	85.4	229	14	US-10-032-221B-2
17	129	85.4	406	9	US-09-925-302-507
18	129	85.4	1689	15	US-10-372-683-8
19	127	84.1	27	14	US-10-032-221B-42
20	125	82.8	22	14	US-10-206-699-266
21	125	82.8	25	14	US-10-032-221B-37
22	117	77.5	25	14	US-10-032-221B-37
23	115	76.2	22	14	US-10-206-699-265
24	115	76.2	1759	15	US-10-369-493-7032
25	113	74.8	22	14	US-10-206-699-267
26	110	72.8	20	14	US-10-206-699-289
27	110	72.8	20	14	US-10-032-221B-29
28	107	70.9	46	9	US-09-864-761-48095
29	107	70.9	1744	15	US-10-369-493-5832
30	105	69.5	142	9	US-09-864-761-38021
31	105	69.5	228	14	US-10-206-699-307
32	104	68.9	18	14	US-10-206-699-254
33	104	68.9	18	14	US-10-206-699-260
34	102	67.5	227	14	US-10-206-699-303
35	102	67.5	227	14	US-10-032-221B-6
36	102	67.5	430	9	US-09-925-302-518
37	102	67.5	459	15	US-10-331-496A-27
38	102	67.5	459	15	US-10-372-683-30
39	102	67.5	1712	10	US-09-961-403-9
40	98	64.9	18	14	US-10-206-699-259
41	98	64.9	22	14	US-10-206-699-270
42	96	63.6	18	14	US-10-206-699-261
43	95	62.9	22	14	US-10-206-699-268
44	94	62.3	18	14	US-10-206-699-253
45	94	62.3	20	14	US-10-206-699-290

#### ALIGNMENTS

#### RESULT 1

US-10-032-221B-39  
; Sequence 39, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T8 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length Tumor  
; OTHER INFORMATION: atin molecule)  
US-10-032-221B-39  
Query Match 100.0%; Score 151; DB 14; Length 27;  
Best Local Similarity 100.0%; Pred. No. 4.8e-16;



CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 20  
TYPE: PRT  
LENGTH: 124  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)  
US-10-032-221B-20

Query Match 96.7%; Score 146; DB 14; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1.4e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 6  
US-10-032-221B-23  
Sequence 23, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 23  
LENGTH: 132  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)  
US-10-032-221B-23

Query Match 96.7%; Score 146; DB 14; Length 132;  
Best Local Similarity 100.0%; Pred. No. 1.5e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
Db 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 7  
US-10-032-221B-22  
Sequence 22, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 22  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)  
US-10-032-221B-22

Query Match 96.7%; Score 146; DB 14; Length 191;  
Best Local Similarity 100.0%; Pred. No. 2.2e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
Db 16 QRFTHMPFLFCNVNDVCFASRNDYS 41

RESULT 8  
US-10-270-877-46  
Sequence 46, Application US/10270877  
Publication No. US20030049791A1  
GENERAL INFORMATION:  
APPLICANT: Saus, Juan  
TITLE OF INVENTION: Goodpasture Binding Protein  
FILE REFERENCE: 98-723-AD1  
CURRENT APPLICATION NUMBER: US/10/270,877  
CURRENT FILING DATE: 2002-10-11  
PRIOR APPLICATION NUMBER: 09/512,563  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/121,483  
PRIOR FILING DATE: 1999-02-24  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 46  
LENGTH: 211  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-877-46

Query Match 96.7%; Score 146; DB 14; Length 211;





; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 40

; LENGTH: 27

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8-3)  
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine  
; OTHER INFORMATION: residues at positions 79 and 85)

US-10-032-221B-40

Query Match

Best Local Similarity 86.8%; Score 131; DB 14; Length 27;

Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

Db 1 KQRTTTPFLFCNVNDVCFASRNDYS 27

RESULT 13

US-10-206-699-306

; Sequence 306, Application US/10206699

; Publication No. US20030100510A1

; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.

; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer

; FILE REFERENCE: MBHB 01-1017

; CURRENT APPLICATION NUMBER: US/10/206,699

; CURRENT FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362

; PRIOR FILING DATE: 2002-06-03

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 306

; LENGTH: 229

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: alpha 5 chain

US-10-206-699-306

Query Match

Best Local Similarity 86.1%; Score 130; DB 14; Length 229;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 55 RRFSTMPFLFCNVNDVCFASRNDYS 80

RESULT 14

US-09-925-297-496

; Sequence 496, Application US/09925297

; Patent No. US20020081659A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: FA105

; CURRENT APPLICATION NUMBER: US/09/925,297

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05989

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 928

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 496

; LENGTH: 309

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (247)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-297-496

Query Match

Best Local Similarity 86.1%; Score 130; DB 9; Length 309;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 135 RRFSTMPFLFCNVNDVCFASRNDYS 160

RESULT 15

US-10-206-699-302

; Sequence 302, Application US/10206699

; Publication No. US20030100510A1

; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.

; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer

; FILE REFERENCE: MBHB 01-1017

; CURRENT APPLICATION NUMBER: US/10/206,699

; CURRENT FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362

; PRIOR FILING DATE: 2002-06-03

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 302

; LENGTH: 229

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: alpha 1 chain

US-10-206-699-302

Query Match

Best Local Similarity 85.4%; Score 129; DB 14; Length 229;

Matches 21; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27

Db 55 RRFSTMPFLFCNVNDVCFASRNDYS 80

Search completed: April 5, 2004, 07:36:06

Job time : 16.9322 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221B-39  
Perfect score: 151  
Sequence: 1 QRFTHMPFLFCNVNDVCFASRNDYS 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/FCUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	146	96.7	211	4	US-09-512-563C-46
2	146	96.7	218	2	US-08-399-889-25
3	146	96.7	218	3	US-09-167-364-25
4	146	96.7	218	3	US-09-439-897-4
5	146	96.7	268	4	US-09-589-927-6
6	146	96.7	268	4	US-09-277-665-6
7	146	96.7	268	4	US-09-589-987-6
8	145	96.0	471	2	US-08-399-889-24
9	145	96.0	471	3	US-09-167-364-24
10	145	96.0	471	3	US-09-439-897-2
11	130	86.1	264	4	US-09-589-927-10
12	130	86.1	264	4	US-09-277-665-10
13	130	86.1	264	4	US-09-589-987-10
14	129	85.4	260	4	US-09-589-927-2
15	129	85.4	260	4	US-09-277-665-2
16	129	85.4	260	4	US-09-589-987-2
17	105	69.5	260	4	US-09-589-927-12
18	105	69.5	260	4	US-09-277-665-12
19	105	69.5	260	4	US-09-589-987-12
20	102	67.5	258	4	US-08-399-889-24
21	102	67.5	258	4	US-09-277-665-4
22	102	67.5	258	4	US-09-589-987-4
23	88	58.3	260	4	US-09-589-927-8
24	88	58.3	260	4	US-09-277-665-8
25	88	58.3	260	4	US-09-589-987-8
26	68	45.0	1694	1	US-08-494-168-2
27	51	33.8	107	3	US-09-102-528-23

28	51	33.8	107	3	US-09-102-528-27	Sequence 27, Appl
29	51	33.8	587	3	US-09-102-528-30	Sequence 30, Appl
30	51	33.8	736	3	US-09-102-528-29	Sequence 29, Appl
31	46	30.5	1117	4	US-09-252-991A-23416	Sequence 23416, A
32	45	29.8	410	4	US-09-543-681A-5407	Sequence 5407, Ap
33	44	29.1	49	1	US-07-865-166A-6	Sequence 6, Appl
34	44	29.1	326	4	US-09-134-000C-4813	Sequence 4813, Ap
35	43.5	28.6	683	4	US-09-194-468A-30	Sequence 30, Appl
36	43	28.5	72	4	US-09-543-681A-8206	Sequence 8206, Ap
37	43	28.5	118	3	US-09-413-814-36	Sequence 36, Appl
38	43	28.5	256	4	US-09-194-146-4	Sequence 4, Appl
39	42.5	28.1	340	1	US-08-118-270-49	Sequence 49, Appl
40	42.5	28.1	340	5	PCT-US93-08528-49	Sequence 49, Appl
41	42.5	28.1	372	1	US-07-937-609-20	Sequence 20, Appl
42	42.5	28.1	372	3	US-08-029-170-20	Sequence 20, Appl
43	42.5	28.1	407	1	US-08-117-965-26	Sequence 26, Appl
44	42.5	28.1	407	2	US-08-390-000A-6	Sequence 6, Appl
45	42.5	28.1	407	5	PCT-US92-06532-3	Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-09-512-563C-46  
; Sequence 46, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512,563C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-09-512-563C-46

Query Match 96.7% Score 146; DB 4; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.6e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFCNVNDVCFASRNDYS 27  
DB 69 QRFTHMPFLFCNVNDVCFASRNDYS 94

RESULT 2  
US-08-399-889-25  
; Sequence 25, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 218  
; TYPE: PRT

```
; ORGANISM: Human
US-09-399-889-25

Query Match
Best Local Similarity 96.7%; Score 146; DB 2; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTTPFLFCNVNDVCFASRNDYS 68

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match
Best Local Similarity 96.7%; Score 146; DB 3; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTTPFLFCNVNDVCFASRNDYS 68

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match
Best Local Similarity 96.7%; Score 146; DB 3; Length 218;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
Db 43 QRTTTPFLFCNVNDVCFASRNDYS 68

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
Db 93 QRTTTPFLFCNVNDVCFASRNDYS 118

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRTTTPFLFCNVNDVCFASRNDYS 27
Db 93 QRTTTPFLFCNVNDVCFASRNDYS 118

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match
Best Local Similarity 96.7%; Score 146; DB 4; Length 268;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



; CURRENT APPLICATION NUMBER: US/09/277,663

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQRTTWPFLFSNVNDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	89.4	220	2	CGU36
2	126	89.4	1670	1	CGHU3B
3	125	88.7	471	2	A39024
4	120	85.1	161	2	S49488
5	120	85.1	246	2	I48302
6	110	78.0	253	2	I48304
7	110	78.0	754	2	A5267
8	110	78.0	1691	1	S22917
9	109	77.3	258	2	B61228
10	109	77.3	1669	1	CGHU4B
11	109	77.3	1869	1	CGMS4B
12	100	70.9	1747	2	A54121
13	100	70.9	1752	2	A45407
14	95	67.4	1758	2	T29350
15	95	67.4	1759	2	T19351
16	95	67.4	1763	2	S16366
17	91	64.5	261	2	A34476
18	87	61.7	1744	2	S40991
19	85	60.3	1691	1	CGHU6B
20	82	58.2	775	2	A61228
21	82	58.2	1707	2	A33526
22	82	58.2	1712	1	CGHU2B
23	69	48.9	1761	2	T13990
24	68	48.2	312	2	I48303
25	68	48.2	623	2	A45137
26	68	48.2	1690	1	CGHU1B
27	68	48.2	1775	2	A31893
28	67	47.5	453	2	S18804
29	52.5	37.2	610	2	C70126

30 48 34.0 332 2 F82140  
31 48 34.0 363 2 T37630  
32 48 34.0 509 2 AD0648  
33 47.5 33.7 334 2 C71718  
34 47 33.3 364 2 A97335  
35 46.5 33.0 334 2 B97715  
36 46.5 33.0 955 2 E84845  
37 46 32.6 155 2 B83124  
38 46 32.6 457 2 T33494  
39 46 32.6 585 2 S15963  
40 45.5 32.3 432 2 S51901  
41 45.5 32.3 490 2 T24497  
42 45 31.9 261 2 B81823  
43 45 31.9 452 2 A25346  
44 45 31.9 472 2 AB0907  
45 45 31.9 623 2 G95180

## ALIGNMENTS

### RESULT 1

B49736

collagen alpha 3(IV) chain, medium splice form - human (fragment)

N;Contains: collagen alpha 3(IV) chain, splice form GP-V

C;Species: Homo sapiens (man)

C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text\_change 17-Mar-2000

C;Accession: B49736; D49736; S69111

R;Peng, L.; Xia Y.; Wilson, C.B

J. Biol. Chem. 269, 2342-2348, 1994

A;Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.

A;Reference number: A49736; MUID:94124597; PMID:8294492

A;Accession: B49736

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 169-220 <FEN1>

A;Accession: D49736

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 22-220 <FEN2>

A;Cross-references: GB:U02519; NID:G409106; PIDN:AAAI8942.1; PID:G409107

A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank

R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wi

Eur. J. Biochem. 229, 754-760, 1995

A;Title: Characterization and expression of multiple alternatively spliced transcripts

utoantigen and one of its alternative forms.

A;Reference number: S69111; MUID:95278230; PMID:7758473

A;Accession: S69111

A;Molecule type: mRNA

A;Residues: 1-45,169-204,'L',206-220 <PEN>

C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.

C;Genetics:

A;Gene: GDB:COL4A3

A;Cross-references: GDB:128351; OMIM:120070

A;Map position: 2q36-2q37

C;Superfamily: collagen alpha 1(IV) chain

F;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrac

F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status prec

F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status

F;22-220/Domain: carboxyl-terminal nonhelical, NC1 <NC1>

F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 89.4%; Score 126; DB 2; Length 220;

Best Local Similarity 92.3%; Pred. No. 1.5e-11;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 QRTTWPFLFSNVNDFASRNDYS 27

Db 78 QRTTWPFLFCNVNDFASRNDYS 103

### RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human  
 N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
 C:Species: Homo sapiens (man)  
 C>Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
 C:Accession: A54763; A43928; A44043; A45971; A39786  
 R:Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Read, S.T.  
 J. Biol. Chem. 269, 23013-23017, 1994  
 A>Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
 A:Reference number: A54763; MUID:94364994; PMID:8083201  
 A:Accession: A54763  
 A:Molecule type: mRNA  
 A:Residues: 1-1670 <NAR>  
 A:CROSS-references: GB:X80031; NID:G577563; PID:G577564  
 A:Experimental source: kidney  
 A:Turner, N.; Mason, P. J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
 J. Clin. Invest. 89, 592-601, 1992  
 A>Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha  
 A:Reference number: A43928; MUID:92147878; PMID:1737849  
 A:Accession: A43928  
 A:Molecule type: mRNA  
 A:Residues: 1331-1524, 'I', 1526-1670 <TUR>  
 A:CROSS-references: GB:X81379  
 A:Experimental source: kidney  
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 267, 19780-19784, 1992  
 A>Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture  
 A:Reference number: A44043; MUID:93015826; PMID:1400291  
 A:Accession: A44043  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1386-1670 <QUI>  
 A:CROSS-references: GB:M92993; NID:G177895; PID:AA21610.1; PID:G177896  
 A>Note: Sequence extracted from NCBI backbone (NCBI:P:115597)  
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 269, 17358, 1994  
 A:Reference number: A44738; MUID:94274734; PMID:8006044  
 A:Contents: annotation; erratum; correction to intronic sequence in A44043  
 R:Bernal, D.; Quinones, S.; Saus, J.  
 J. Biol. Chem. 268, 12090-12094, 1993  
 A>Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
 A:Reference number: A45971; MUID:93280184; PMID:8505332  
 A:Accession: A45971  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1427-1444 <BER>  
 A>Note: Sequence extracted from NCBI backbone (NCBI:P:133363); sequence incorrectly ident  
 R:Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Read, S.T.  
 Am. J. Hum. Genet. 49, 545-554, 1991  
 A>Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
 A:Reference number: A39786; MUID:91353570; PMID:1882840  
 A:Accession: A39786  
 A:Molecule type: mRNA  
 A:Residues: 1453-1593, 'A', 1595-1670 <MOR>  
 A:CROSS-references: GB:S55790; NID:G234418; PID:AA19637.1; PID:G234419  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
 C:Genetics:  
 A:Gene: GDB:COL4A3  
 A:CROSS-references: GDB:128351; OMIM:120070  
 A:Map position: 2q36-q37  
 A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
 A>Note: The alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
 C:Function:  
 A:Description: minor structural component of extracellular basement membrane in kidney g  
 C:Superfamily: collagen alpha 1(IV) chain  
 C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
 F1-29/Domain: signal sequence #status predicted <SIG>  
 F129-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <NAT>  
 F129-42/Domain: amino-terminal nonhelical, NH1 <NH1>

F143-1438/Region: interrupted helical  
 F1791-793/Region: cell attachment (R-G-D) motif  
 F1996-998/Region: cell attachment (R-G-D) motif  
 F1154-1156/Region: cell attachment (R-G-D) motif  
 F1306-1308/Region: cell attachment (R-G-D) motif  
 F1345-1347/Region: cell attachment (R-G-D) motif  
 F1432-1434/Region: cell attachment (R-G-D) motif  
 F1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
 F1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
 F1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
 F13133, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: interchain #status pred  
 F1253/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F1505-1511, 1616-1622/Disulfide bonds: #status predicted  
 F1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
 Query Match 89.4%; Score 126; DB 1; Length 1670;  
 Best Local Similarity 92.3%; Pred. No. 1.5e-10;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 QRTTMTPELFNSVNDVSNFASRNDYS 27  
 |||||  
 DB 1495 QRTTMTPELFNSVNDVSNFASRNDYS 1520  
 |||||  
 RESULT 3  
 A39024  
 collagen alpha 3(IV) chain - bovine (fragment)  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 04-Dec-1992 #sequence revision 04-Dec-1992 #text change 13-Aug-1999  
 C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815  
 R:Morrison, K.E.; Germino, G.G.; Read, S.T.  
 J. Biol. Chem. 266, 34-39, 1991  
 A>Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
 A:Reference number: A39024; MUID:91093146; PMID:1985905  
 A:Accession: A39024  
 A:Molecule type: mRNA  
 A:Residues: 1-471 <NOR>  
 A:CROSS-references: EXBL:M63139; NID:G162886; PID:AA62708.1; PID:G162887  
 R:Butkowsky, R.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
 J. Biol. Chem. 262, 7874-7877, 1987  
 A>Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
 A:Reference number: S18432; MUID:87222419; PMID:2438283  
 A:Accession: S20672  
 A:Molecule type: protein  
 A:Residues: 227-228, 'X', 230-244 <BUT>  
 R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
 J. Biol. Chem. 263, 13374-13380, 1988  
 A>Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
 A:Reference number: S17802; MUID:88330844; PMID:3417661  
 A:Accession: S17802  
 A:Molecule type: protein  
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
 R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
 J. Biol. Chem. 265, 5466-5469, 1990  
 A>Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
 A:Reference number: A35167; MUID:90202779; PMID:2318822  
 A:Accession: A35167  
 A:Molecule type: protein  
 A:Residues: 236-255 <GUN>  
 R:Gunwar, S.; Ballerster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; No  
 J. Biol. Chem. 266, 15318-15324, 1991  
 A>Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol  
 A:Reference number: A39419; MUID:91332055; PMID:1869555  
 A:Accession: C39419  
 A:Molecule type: protein  
 A:Residues: 236-255 <GUT>  
 C:Superfamily: collagen alpha 1(IV) chain  
 C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;  
 F1-238/Domain: collagenous (fragment) #status predicted <COL>  
 F129-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
 F129-353/Domain: repeat NC1 #status predicted <NC1>  
 F1354-471/Domain: repeat NC1 #status predicted <NC12>



F:232,238/Modified site: hydroxyproline (Pro) #status experimental  
F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 88.7%; Score 125; DB 2; Length 471;  
Best Local Similarity 88.5%; Pred. No. 5.1e-11;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
DB 296 QRFTHMPFLFCNVNDVSNFASRNDYS 321

RESULT 4  
S49488  
collagen alpha 3(IV) chain - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: S49488  
R:Oberbaeumer, I.  
submitted to the EMBL Data Library, October 1994  
A:Description: Cloning of the NCI domains fo the minor collagen IV chains of mouse via F  
ells.  
A:Reference number: S49487  
A:Accession: S49488  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-161 <OBE>  
A:Cross-references: EMBL:X82205; NID:G559472; PIDN:CAA57689.1; PID:G559916  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 85.1%; Score 120; DB 2; Length 161;  
Best Local Similarity 84.6%; Pred. No. 8.4e-11;  
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
DB 4 QRFTHMPFLFCNVNDVSNFASRNDYS 29

RESULT 5  
I48302  
collagen alpha 3(IV) chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
C:Accession: I48302; S47278  
R:Miner, J.H.; Sanes, J.R.  
J. Cell Biol. 127, 879-891, 1994  
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
A:Reference number: A54979; MUID:95050957; PMID:7962065  
A:Accession: I48302  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-246 <RES>  
A:Cross-references: EMBL:Z35166; NID:G535197; PID:G535198  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 85.1%; Score 120; DB 2; Length 246;  
Best Local Similarity 84.6%; Pred. No. 1.4e-10;  
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
DB 71 QRFTHMPFLFCNVNDVSNFASRNDYS 96

RESULT 6  
I48304  
collagen alpha 5(IV) chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999  
C:Accession: I48304; S47280  
R:Miner, J.H.; Sanes, J.R.  
J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq  
A:Reference number: A54979; MUID:95050957; PMID:7962065  
A:Accession: I48304  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <RES>  
A:Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 78.0%; Score 110; DB 2; Length 253;  
Best Local Similarity 73.1%; Pred. No. 4.6e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
DB 79 RRFSTWPFNFCNVNDVSNFASRNDYS 104

RESULT 7  
A55267  
collagen alpha 5(IV) chain - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: A55267  
R:Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994  
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-J  
en type IV.  
A:Reference number: A55267; MUID:94224868; PMID:8171024  
A:Accession: A55267  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-754 <ZHE>  
A:Cross-references: GB:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 78.0%; Score 110; DB 2; Length 754;  
Best Local Similarity 73.1%; Pred. No. 1.6e-08;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
DB 587 RRFSTWPFNFCNVNDVSNFASRNDYS 612

RESULT 8  
S22917  
collagen alpha 5(IV) chain precursor, renal splice form - human  
N:Alternate names: procollagen alpha 5(IV) chain  
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 27-Feb-1997 #text\_change 21-Jul-2000  
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A3  
R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident  
n Alport syndrome patient.  
A:Reference number: S22917; MUID:92316923; PMID:1352287  
A:Accession: S22917  
A:Molecule type: mRNA  
A:Residues: 1-967 <ZHO>  
A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234  
R:Zhou, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994

A:Title: Structure of the human type IV collagen COL4A5 gene.  
A:Reference number: A54365; MUID:94165049; PMID:8120014  
A:Accession: A54365  
A:Molecule type: DNA  
A:Residues: 1-922 <ZHO>  
A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AAC27816.1; P  
R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggva  
Science 261, 1167-1169, 1993  
A:Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited s

[illegible]

D5 84 RKFTMPFLFCNNVNCVASFASNDYS 109

## RESULT 10

COLH4B  
collagen alpha 1(IV) chain precursor - human  
N:Alternate names: procollagen alpha 1(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C:Accession: S16876; A32117; S02738; S00048; S28826; A23115; S0207; S39614; A02863; A58  
R:Solinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen  
A:Reference number: S16876; MUID:89340433; PMID:2701944  
A:Accession: S16876  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1669 <SO11>  
A:Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R:Solinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A:Reference number: A32690; MUID:89034231; PMID:3182844  
A:Accession: A32117  
A:Molecule type: DNA  
A:Residues: 1-28 <SO12>  
A:Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R:Poeschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane  
A:Reference number: S02738; MUID:89030632; PMID:2846280  
A:Accession: S02738  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-6, 'L', 8-28 <POE>  
A:Cross-references: EMBL:X12784; NID:G30072  
R:Bräzel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;  
Eur. J. Biochem. 158, 529-536, 1987  
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem  
A:Reference number: S00048; MUID:88029471; PMID:3311751  
A:Accession: S00048  
A:Molecule type: mRNA  
A:Residues: 1-318, 'A', 320-944 <BRAL>  
A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A:Accession: S25826  
A:Molecule type: protein  
R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (75  
A:Reference number: A23115; MUID:86004708; PMID:4043082  
A:Accession: A23115  
A:Molecule type: protein  
A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A:Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
R:EBle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen  
A:Reference number: S39614; MUID:94038963; PMID:8223488  
A:Accession: S39614  
A:Molecule type: protein  
A:Residues: 371-554 <ESU>  
R:Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984  
A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid  
A:Reference number: A02863; MUID:85003629; PMID:6434307  
A:Accession: A02863  
A:Molecule type: protein  
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'X', 999;  
A:Experimental source: placenta  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A:Title: Peptin fragments of human placental basement-membrane collagens showing inter  
A:Reference number: S16908; MUID:82005835; PMID:6792033  
A:Accession: A58517  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-1  
R:MacWhight, R.S.; Benson, V.A.; Lovellio, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane  
A:Reference number: S16910; MUID:84053346; PMID:6416291  
A:Accession: S16910  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948;  
A:Experimental source: placenta  
R:Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;  
J. Biol. Chem. 260, 7681-7687, 1985  
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen  
A:Reference number: S01466; MUID:85207819; PMID:2581969  
A:Accession: S01466  
A:Molecule type: mRNA  
A:Residues: 1256-1669 <PIH>  
A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424  
R:Brinker, J.M.; Gudus, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains  
A:Reference number: S16879; MUID:85216555; PMID:2582422  
A:Accession: S16879  
A:Molecule type: mRNA  
A:Residues: 1259-1669 <BRI>  
A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818  
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,  
Eur. J. Biochem. 147, 217-224, 1985  
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-  
A:Reference number: A02864; MUID:85127033; PMID:2578961  
A:Accession: S19091  
A:Molecule type: protein  
A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',  
R:Siebold, B.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 176, 617-624, 1988  
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyter  
A:Reference number: S02550; MUID:89005112; PMID:2844531  
A:Contents: annotation; disulfide bonds  
C:Genetics:  
A:Gene: GDB:COL4A1  
A:Cross-references: GDB:119791; OMIM:120130  
A:Map position: 13q34-13q34  
A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231  
/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 1020/1; 1066/3; 1109/1; 1136/1; 11  
C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2  
C:Associations among trimer amino-terminal domains (disulfide and desmosine cross-links), di  
C:Trimer associations in the interrupted helical domain (with disulfide and desmosine c  
C:Function:  
A:Description: structural component of extracellular basement membrane  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicat  
P:1-26/Domain: signal sequence #status predicted <SIG>  
P:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
P:163-1440/Domain: amino-terminal nonhelical, 78 <7SD>  
P:163-1440/Domain: interrupted helical <COL>  
P:29-162/Domain: amino-terminal nonhelical, 78 <7SD>  
F:414-452/Region: integrin binding #status experimental  
F:597-599/Region: cell attachment (R-G-D) motif  
F:917-919/Region: cell attachment (R-G-D) motif  
F:968-970/Region: cell attachment (R-G-D) motif  
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <C11>

R.;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.  
Gene 43, 301-304, 1986  
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeoxynucleotide as a probe.  
A;Reference number: A25636; PMID:86301886; PMID:3755692  
A;Accession: A25636  
A;Molecule type: mRNA  
A;Residues: 1149-1396, 'S', 1398-1424 <NAT>  
A;Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287  
A;Note: The authors translated the codon CAG for residue 1374 as Arg.  
A;Author: Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla  
R.;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla  
J. Biol. Chem. 262, 8496-8499, 1987  
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
collagen and human alpha-1(IV) collagen.  
A;Reference number: A94680; PMID:87250460; PMID:3597383  
A;Accession: A29301  
A;Molecule type: mRNA  
A;Residues: 1441-1659 <EUR>  
A;Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115  
A;Author: Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla  
J. Biol. Chem. 263, 12310-12314, 1988  
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequen  
A;Reference number: S19079; PMID:88315019; PMID:2842328  
A;Accession: S19079  
A;Molecule type: DNA  
A;Residues: 1-28 <X2>  
A;Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503  
A;Author: Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A;Title: Head-to-head arrangement of murine type IV collagen genes.  
A;Reference number: A94220; PMID:89071759; PMID:3198626  
A;Accession: A32003  
A;Molecule type: DNA  
A;Residues: 1-28 <XAY>  
A;Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449  
A;Author: Burdello, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
A;Reference number: A94220; PMID:89071759; PMID:3200851  
A;Accession: A31766  
A;Molecule type: DNA  
A;Residues: 1-28 <EUR>  
A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668  
A;Author: Sakurai, Y.; Sullivan, M.; Yamada, Y.  
J. Biol. Chem. 261, 6654-6657, 1986  
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen gene  
A;Reference number: S19094; PMID:86196099; PMID:3009468  
A;Accession: S19094  
A;Molecule type: DNA  
A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>  
A;Cross-references: EMBL:M3027  
A;Author: Ruchpattan, D.; Timpl, R.; Glanville, R.W.  
FEBS Lett. 115, 297-300, 1980  
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (t  
A;Reference number: S16909; PMID:80246483; PMID:6772473  
A;Accession: S16909  
A;Molecule type: protein  
A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-12  
A;Cross-references: EMBL:M3027  
A;Author: Ruchpattan, D.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 123, 505-512, 1982  
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial ami  
A;Reference number: A25991; PMID:82186723; PMID:6804236  
A;Accession: A25991  
A;Molecule type: protein  
A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 10  
61, 'X', 1063-1085, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119  
A;Accession: B2591  
A;Molecule type: protein  
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234  
17, '3', 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DK', 1275, 'I', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 132  
R.;Reber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 139, 401-410, 1984  
A;Title: Subunit structure and assembly of the globular domain of basement-membrane col  
A;Reference number: S17801; PMID:84132058; PMID:6698021  
A;Accession: S17801

A:Molecule type: protein  
A:Residues: 1435-1443 <WEB>

C:Genetics:

A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3

A>Note: the list of introns may be incomplete

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>

F:28-162/Domain: 7S <7SD>

F:163-1440/Domain: collagenous, triple helix <COL>

F:597-599/Region: cell attachment (R-G-D) motif

F:781-783/Region: cell attachment (R-G-D) motif

F:917-919/Region: cell attachment (R-G-D) motif

F:968-970/Region: cell attachment (R-G-D) motif

F:1441-1569/Domain: carboxyl-terminal nonhelical, NCI <NCI>

F:1441-1552/Region: duplication

F:1553-1569/Region: duplication

F:31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted

F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:971.974.977.986.989.1001.1007.1019.1022.1031.1037.1040.1055.1060.1063.1075.1078.1090.1

92.1298.1310.1313.1322.1337.1346.1349.1422.1425.1431.1437.1440/Modified site: hydroxypro

F:1214.1424/Modified site: 4-hydroxyproline (Pro) #status experimental

F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental

F:1503-1511.1516-1522/Disulfide bonds: #status predicted

Query Match 77.3%; Score 109; DB 1; Length 1669;

Best Local Similarity 73.1%; Pred. No. 5.7e-08;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFSTMPFLFSNVNVSFASRNDYS 27

DB 1495 RRFSTMPFLFCNNVNCVAFSNDYS 1520

RESULT 12

A54121

collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)

N:Alternate names: collagen alpha 2(IV) chain homolog

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 13-Aug-1999

A:Accession: A54121; S43317

R:Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.

J. Biol. Chem. 269, 13167-13171, 1994

A:Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I

A:Reference number: A54121; MUID:94230414; PMID:8175744

A:Molecule type: mRNA

A:Accession: A54121

A:Residues: 1-1747 <EXP>

A:Cross-references: EMBL:X76730; NID:g483606; PIDN:CAA54146.1; PID:g483607

C:Genetics:

A:Gene: COLP4alpha

C:Superfamily: collagen alpha 1(IV) chain

Query Match

Best Local Similarity 70.9%; Score 100; DB 2; Length 1747;

Matches 19; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFSTMPFLFSNVNVSFASRNDYS 27

DB 1576 QRFSTMPFLFCNNVNCVAFSNDYS 1601

RESULT 13

A5407

collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999

A:Accession: A5407; A43903; A23940

R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.

J. Biol. Chem. 268, 5249-5254, 1993

A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and ana

A:Reference number: A5407; MUID:93186842; PMID:844899

A:Accession: A45407

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1752 <EXP>

A>Note: sequence extracted from NCBI backbone (NCBIP:126841)

R:Wessel, G.M.; Etkin, M.; Benson, S.

Dev. Biol. 148, 261-272, 1991

A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously proc

A:Reference number: A43903; MUID:92038439; PMID:1936564

A:Accession: A43903

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 'P', 633-1537, 'G' <VES>

A:Cross-references: GB:S64572; NID:9238616; PIDN:AAB20270.1; PID:9238617

A>Note: sequence extracted from NCBI backbone (NCBIN:64572, NCBIP:64573)

R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.

Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986

A:Title: Structure and developmentally regulated expression of a Strongylocentrotus pur

A:Reference number: A23940; MUID:86205894; PMID:3458186

A:Accession: A23940

A:Molecule type: DNA

A:Residues: 742-812 <VEN>

A:Cross-references: EMBL:M13206

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

F:162-1523/Region: amino-terminal nonhelical, 7S <7SD>

F:1524-1752/Domain: carboxyl-terminal nonhelical, NCI <NCI>

F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>

F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>

F:129/Modified site: allylsine (Lys) #status predicted

Query Match 70.9%; Score 100; DB 2; Length 1752;

Best Local Similarity 69.2%; Pred. No. 1.4e-06;

Matches 18; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRFSTMPFLFSNVNVSFASRNDYS 27

DB 1578 RRFSTMPFLFCNNVNCVAFSNDYS 1603

RESULT 14

T29350

hypothetical protein F01G12.5a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

A:Accession: T29350

R:Wu, X.; Le, T.T.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F01G12.

A:Reference number: Z20611

A:Accession: T29350

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1758 <WUX>

A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a

A:Experimental source: strain Bristol N2; clone F01G12

C:Genetics:

A:Gene: CESP:F01G12.5a

A:Map position: X

A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/

C:Superfamily: collagen alpha 1(IV) chain

Query Match

Best Local Similarity 67.4%; Score 95; DB 2; Length 1758;

Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFSTMPFLFSNVNVSFASRNDYS 27

DB 1581 QRFSTMPFLFCNNVNCVAFSNDYS 1606

RESULT 15

Mon Apr 5 07:53:14 2004

T29351  
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans  
N/Alternate names: collagen alpha 2(IV) chain precursor c1b-1  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C/Accession: T29351  
R/Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A/Description: The sequence of C. elegans cosmid F01G12.  
A/Reference number: Z20611  
A/Accession: T29351  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-1759 <WUX>  
A/Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a  
A/Experimental source: strain Bristol N2, clone F01G12  
C/Genetics:  
A/Gene: CESP:F01G12.5a  
A/Map position: X  
A/Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3  
C/Superfamily: collagen alpha 1(IV) chain

Query Match 67.4%; Score 95; DB 2; Length 1759;  
Best Local Similarity 69.2%; Pred. No. 8e-06;  
Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRFTTPELFSNVNDVSNFASRNDYS 27  
|||:|||||:|:|:|  
Db 1582 QRFTTPELFCDFNNVCNYSRNDKS 1607

Search completed: April 5, 2004, 07:05:38  
Job time : 6.5569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.39952 Seconds

(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQRTTWFPLFSNVNDVSNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	89.4	1670	1 CA34_HUMAN	Q01955 homo sapien
2	125	88.7	471	1 CA34_BOVIN	Q28084 bos taurus
3	110	78.0	754	1 CA54_CANFA	Q28247 canis famil
4	110	78.0	1685	1 CA54_HUMAN	P29400 homo sapien
5	109	77.3	1669	1 CA14_HUMAN	P02462 homo sapien
6	109	77.3	1669	1 CA14_MOUSE	P02463 mus musculus
7	95	67.4	1763	1 CA24_ASCSU	P27393 ascaris suu
8	91	64.5	1758	1 CA24_CAEEL	P17140 caenorhabdi
9	87	61.7	1758	1 CA14_CAEEL	P17139 caenorhabdi
10	85	60.3	1691	1 CA64_HUMAN	Q14031 homo sapien
11	82	58.2	1707	1 CA24_MOUSE	P08122 mus musculus
12	82	58.2	1712	1 CA24_HUMAN	P08572 homo sapien
13	68	48.2	623	1 CA44_RABIT	P55787 oryctolagus
14	68	48.2	1690	1 CA14_HUMAN	P53420 homo sapien
15	68	48.2	1775	1 CA14_DROME	P08120 drosophila
16	67	47.5	453	1 CA44_BOVIN	Q29442 bos taurus
17	52.5	37.2	610	1 MUTL_BORBU	O51229 borrelia bu
18	47.5	33.7	334	1 Y032_RICPR	Q92655 rickettsia
19	46	32.6	457	1 PH4L_CAEEL	P09325 caenorhabdi
20	45.5	32.3	432	1 SYWC_YEAST	Q12109 saccharomyc
21	45	31.9	452	1 CN17_DICDI	P12019 dictyosteli
22	44	31.2	172	1 Y427_UREUR	Q56564 ureaplasma
23	44	31.2	371	1 CYB_ELANI	Q9mlk8 elapsoides
24	44	31.2	664	1 MS16_YEAST	P15424 saccharomyc
25	44	31.2	772	1 C1PB_CLOTM	Q01866 clostridium
26	44	31.2	793	1 MUTS_THERMA	P74926 thermotoga
27	44	31.2	841	1 MYFC_YEREN	P33408 yersinia en
28	44	31.2	1853	1 C1PA_CLOTM	Q06851 clostridium
29	43.5	30.9	288	1 T2D2_STRPN	P09357 streptococc
30	43	30.5	145	1 HA17_CLOBO	P46083 clostridium
31	43	30.5	282	1 PANC_AQUAE	O67891 aquifex aeo
32	43	30.5	370	1 CYB_MICKI	Q9mlk2 micropechis
33	43	30.5	371	1 CYB_LOXBI	O48100 loxocemus b

## RESULT 1

CA34\_HUMAN STANDARD; PRT; 1670 AA.  
ID CA34\_HUMAN  
AC Q01955; Q9BOT2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94364994; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RL J. Biol. Chem. 269:23013-23017(1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;  
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND  
RP CVS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;  
RP PRO-574; GLU-1269 AND PRO-1474.  
RX MEDLINE=21064696; PubMed=1134255;  
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
RN [4]  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:19780-19784(1992).  
RN [5]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Readers S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554(1991).  
RN [6]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;

## ALIGNMENTS

34	43	30.5	397	1	YMP7_YEAST	Q04359 saccharomyc
35	43	30.5	642	1	YB9W_YEAST	P38352 saccharomyc
36	43	30.5	680	1	PRP2_STRPN	P10524 streptococc
37	43	30.5	1462	1	NCO2_MOUSE	O61026 mus musculu
38	43	30.5	1464	1	NCO2_HUMAN	Q15596 homo sapien
39	43	30.5	1465	1	NCO2_RAT	Q9wll9 rattus norv
40	42	29.8	286	1	VNS4_RSVN	P30847 rice stripe
41	42	29.8	286	1	VNS4_RSVT	Q34961 rice stripe
42	42	29.8	370	1	CYB_EUNFA	Q9mlj8 burgatus fa
43	42	29.8	385	1	CHEE_BORBU	Q45047 borrelia bu
44	42	29.8	385	1	GUNF_FUSOX	P46239 fusarium ox
45	42	29.8	417	1	YGA4_YEAST	P53196 saccharomyc



Mon Apr 5 07:53:14 2004

RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT to be the alpha 3 chain of type IV collagen.";  
RL J. Clin. Invest. 89:592-601(1992).  
RN [7].  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RA Ding J.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [8].  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NCI domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions.";  
RL J. Biol. Chem. 269:2342-2348(1994).  
RN [9].  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Nimomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q35.";  
RL FEBS Lett. 424:11-16(1998).  
RN [10].  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced.";  
RL J. Biol. Chem. 268:12090-12094(1993).  
RN [11].  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemnick H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Monens L.A.H., van Oost B.A., Brunner H.G.,  
RA Redders S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome.";  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=3;  
CC Comment-Additional isoforms seem to exist. Isoforms differ in  
CC the C-terminal part of the NCI domain;  
CC Name=1;  
CC IsoId=Q01955-1; Sequence=Displayed;  
CC Name=2; Synonyms=V;  
CC IsoId=Q01955-2; Sequence=VSP\_001170;  
CC Name=3; Synonyms=L5;  
CC IsoId=Q01955-3; Sequence=VSP\_001171;  
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
CC colocalized and present only in basement membranes of kidney, eye,  
CC cochlea, lung and brain.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NCI) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NCI domain, are conserved in all known type  
CC IV collagens.  
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
CC -!- DISEASE: Antibodies against the NCI domain of alpha3(IV) mediate  
CC the autoimmune disease Goodpasture syndrome [MIM:233450], which is  
CC characterized by hematuria and pulmonary hemorrhage.  
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive  
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder  
CC characterized by progressive glomerulonephritis, renal failure,  
CC hematuria, ocular abnormalities and deafness. The recessive form  
CC occurs equally between males and females.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X80031; CAA56335.1; -!  
CC EMBL; AJ288487; CAC36101.1; -!  
CC EMBL; AJ288488; CAC36101.1; JOINED.  
CC EMBL; AJ288489; CAC36101.1; JOINED.  
CC EMBL; AJ288490; CAC36101.1; JOINED.  
CC EMBL; AJ288491; CAC36101.1; JOINED.  
CC EMBL; AJ288492; CAC36101.1; JOINED.  
CC EMBL; AJ288493; CAC36101.1; JOINED.  
CC EMBL; AJ288494; CAC36101.1; JOINED.  
CC EMBL; AJ288495; CAC36101.1; JOINED.  
CC EMBL; AJ288496; CAC36101.1; JOINED.  
CC EMBL; AJ288497; CAC36101.1; JOINED.  
CC EMBL; AJ288498; CAC36101.1; JOINED.  
CC EMBL; AJ288499; CAC36101.1; JOINED.  
CC EMBL; AJ288500; CAC36101.1; JOINED.  
CC EMBL; AJ288501; CAC36101.1; JOINED.  
CC EMBL; AJ288502; CAC36101.1; JOINED.  
CC EMBL; AJ288503; CAC36101.1; JOINED.  
CC EMBL; AJ288504; CAC36101.1; JOINED.  
CC EMBL; AJ288505; CAC36101.1; JOINED.  
CC EMBL; AJ288506; CAC36101.1; JOINED.  
CC EMBL; AJ288507; CAC36101.1; JOINED.  
CC EMBL; AJ288508; CAC36101.1; JOINED.  
CC EMBL; AJ288509; CAC36101.1; JOINED.  
CC EMBL; AJ288510; CAC36101.1; JOINED.  
CC EMBL; AJ288511; CAC36101.1; JOINED.  
CC EMBL; AJ288512; CAC36101.1; JOINED.  
CC EMBL; AJ288513; CAC36101.1; JOINED.  
CC EMBL; AJ288514; CAC36101.1; JOINED.  
CC EMBL; AJ288515; CAC36101.1; JOINED.  
CC EMBL; AJ288516; CAC36101.1; JOINED.  
CC EMBL; AJ288517; CAC36101.1; JOINED.  
CC EMBL; AJ288518; CAC36101.1; JOINED.  
CC EMBL; AJ288519; CAC36101.1; JOINED.  
CC EMBL; AJ288520; CAC36101.1; JOINED.  
CC EMBL; AJ288521; CAC36101.1; JOINED.  
CC EMBL; AJ288522; CAC36101.1; JOINED.  
CC EMBL; AJ288523; CAC36101.1; JOINED.  
CC EMBL; AJ288524; CAC36101.1; JOINED.  
CC EMBL; AJ288525; CAC36101.1; JOINED.  
CC EMBL; AJ288526; CAC36101.1; JOINED.  
CC EMBL; AJ288527; CAC36101.1; JOINED.  
CC EMBL; AJ288528; CAC36101.1; JOINED.  
CC EMBL; AJ288529; CAC36101.1; JOINED.  
CC EMBL; AJ288530; CAC36101.1; JOINED.  
CC EMBL; AJ288531; CAC36101.1; JOINED.  
CC EMBL; AJ288532; CAC36101.1; JOINED.  
CC EMBL; AJ288533; CAC36101.1; JOINED.  
CC EMBL; AJ288534; CAC36101.1; JOINED.  
CC EMBL; AJ288535; CAC36101.1; JOINED.  
CC EMBL; AJ288536; CAC36101.1; JOINED.

[illegible]

CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (potential).  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC x-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of  
CC canine X-linked hereditary nephritis (HN), a disease similar to  
CC that in humans (also referred to as Alport syndrome) characterized  
CC by progressive renal failure and neurosensory deafness.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U07888; AAB60258.1; --  
CC PIR: A55267; A55267.  
CC InterPro: IPR008161; C1g helix.  
CC InterPro: IPR008160; Collagen.  
CC InterPro: IPR001442; Procollagen4\_C.  
CC Pfam: PF01413; C4; 2.  
CC Pfam: PF01391; Collagen; 8.  
CC ProDom: PDOC0007; C1g helix; 1.  
CC ProDom: PDOC0923; Procollagen4; 1.  
CC SMART: SM00111; C4; 2.  
CC ExTracellul matrix; Connective tissue; Repeat; Hydroxylation;  
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
CC NON\_TER 1  
CC DOMAIN <1 530 TRIPLE-HELICAL REGION.  
CC 531 >754 NON-HELICAL REGION (NC1).  
CC DISULFID 552 643 OR 640 (BY SIMILARITY).  
CC DISULFID 595 640 OR 643 (BY SIMILARITY).  
CC DISULFID 597 603 BY SIMILARITY.  
CC DISULFID 662 ? OR 754 (BY SIMILARITY).  
CC DISULFID 696 -754 BY SIMILARITY.  
CC DISULFID 708 714 BY SIMILARITY.  
CC NON\_TER 754 754  
CC SEQUENCE 754 AA; 73537 MW; DSE321C287FA925B CRC64;  
Query Match 78.0%; Score 110; DB 1; Length 754;  
Best Local Similarity 73.1%; Pred. No. 1.3e-08;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFETMPLEFSNVNDVSNFASNDYS 27  
DB 587 RRFSTMPFMFCNNVNCVNFASNDYS 612  
RESULT 4  
ID CA54 HUMAN STANDARD; PRT; 1685 AA.  
AC P29400; Q16005; Q16126;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 5(IV) chain precursor.  
GN COL4A5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94165049; PubMed=8120014;  
Zhou J., Leinonen A., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A5 gene.";  
J. Biol. Chem. 269:6608-6614(1994).  
RN [2]  
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
RC TISSUE=Kidney;  
RX MEDLINE=923115923; PubMed=1352287;  
Zhou J., Herrz J.M., Leinonen A., Tryggvason K.;  
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen  
chain and identification of a single-base mutation in exon 23  
converting glycine 521 in the collagenous domain to cysteine in an  
Alport syndrome patient.";  
J. Biol. Chem. 267:12475-12481(1992).  
RN [3]  
RP SEQUENCE OF 85-1685 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=90337990; PubMed=2380186;  
Pihlajaniemi T., Pihlajainen E.R., Myers J.C.;  
RT "Complete primary structure of the triple-helical region and the  
carboxyl-terminal domain of a new type IV collagen chain, alpha  
5(IV).";  
J. Biol. Chem. 265:13758-13766(1990).  
RN [4]  
RP SEQUENCE OF 924-1685 FROM N.A.  
RX MEDLINE=91169491; PubMed=2004755;  
Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
RT "Characterization of the 3' half of the human type IV collagen alpha  
5 gene that is affected in the Alport syndrome.";  
Genomics 9:1-9(1991).  
RN [5]  
RP SEQUENCE OF 914-1685 FROM N.A.  
RX MEDLINE=90160375; PubMed=1689491;  
Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtye M., Shows T.B.,  
Tryggvason K.;  
RT "Identification of a distinct type IV collagen alpha chain with  
restricted kidney distribution and assignment of its gene to the  
locus of X chromosome-linked Alport syndrome.";  
Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RP SEQUENCE OF 1442-1471 FROM N.A.  
RX MEDLINE=90252791; PubMed=2339699;  
Myers J.C., Jones T.A., Pihlajainen E.R., Kadri A.S., Goddard A.D.,  
Sheer D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
to the region of the X chromosome containing the Alport syndrome  
locus.";  
Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RP SEQUENCE OF 1-20 FROM N.A.  
RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
Marynen P.;  
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1358-1270 FROM N.A. (ISOFORM 2).  
RX MEDLINE=94133540; PubMed=8301933;  
Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
cells: a complex mutation in the COL4A5 gene of an Alport patient  
deletes the NC1 domain.";  
Kidney Int. 44:1316-1321(1993).  
RN [9]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
Lemlink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
Hum. Mutat. 9:477-499(1997).  
RN [10]  
RP VARIANT AS SER-1564.

RA MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hosikaka S.L., Gregory M.C., Atkin C.L.,  
RA Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome.";  
RL Genomics 9:10-18(1991).  
RN [11]  
RP VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
RA Tryggvason K., Gubler M.-C., Antignac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
RT (IV) chain associated with X-linked Alport syndrome: characterization  
RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
RT cDNA fragments.";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [12]  
RP VARIANT AS ARG-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
RA Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome.";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [13]  
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=9406498;  
RA Lemnick H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
RA Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
RA van Oost B.A., Smeets H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of  
RT patients with Alport syndrome.";  
RL Genomics 17:485-489(1993).  
RN [14]  
RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;  
RX ARG-865; ARG-872 AND CYS-1241.  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States.";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [15]  
RP VARIANT AS ARG-1649.  
RX MEDLINE=96213750; PubMed=8651292;  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States.";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [16]  
RP VARIANTS AS.  
RX MEDLINE=96213754; PubMed=8651296;  
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
RA Savi M., Ballabio A., de Marchi M.;  
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
RT exons of the COL4A5 gene.";  
RL Am. J. Hum. Genet. 58:1192-1204(1996).  
RN [17]  
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
RX MET-1428.  
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,  
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
RA Gubler M.-C., Antignac C.;  
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
RT syndrome.";  
RL Am. J. Hum. Genet. 59:1221-1232(1996).  
RN [18]  
RP VARIANT AS ASP-1498.  
RX MEDLINE=96233932; PubMed=8829632;  
RA Tverskaya S., Bobryna V., Tsalykova F., Ignatova M.,

RA Krasnopolskaya X., Evgrafov O.;  
RT "Substitution of Ala98D in noncollagen domain of alpha 5(IV) collagen  
RT chain associated with adult-onset X-linked Alport syndrome.";  
RL Hum. Mutat. 7:149-150(1996).  
RN [19]  
RP VARIANT AS GLN-1677.  
RX MEDLINE=92795089; PubMed=9150741;  
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
RT "Common ancestry of three Ashkenazi-American families with Alport  
RT syndrome and COL4A5 R1677Q.";  
RL Hum. Genet. 99:681-684(1997).  
RN [20]  
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
RX AND ASP-1536; PubMed=9452056;  
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
RA Trivelli A., Pignatelli A.R., Ragaio M., Rizzoni G.F., de Marchi M.;  
RT "Missense mutations in the COL4A5 gene in patients with X-linked  
RT Alport syndrome.";  
RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
RN [21]  
RP VARIANTS AS VAL-420; 456-PRO-PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
RX 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
RA ARG-1196; GLU-1261; SER-1357 AND ARG-1649.  
RX MEDLINE=99063529; PubMed=9948783;  
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,  
RA Springate J., Shows T.B., Fettereson E., Tryggvason K.;  
RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
RT Alport syndrome using PCR and direct DNA sequencing.";  
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
RN [22]  
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
RX SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
RA MEDLINE=20030197; PubMed=10561141;  
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
RT "Detection of mutations in the COL4A5 gene in over 90% of male  
RT patients with X-linked Alport's syndrome by RT-PCR and direct  
RT sequencing.";  
RL Am. J. Kidney Dis. 34:854-862(1999).  
RN [23]  
RP VARIANT AS ARG-822.  
RX Query Match 78.0%; Score 110; DB 1; Length 1685;  
Best Local Similarity 73.1%; Pred. No. 3.2e-08;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFTHMPFLFSNVNDVSNFASRNDYS 27  
Db 1511 RRFSTMPFMCNINNVCFASRNDYS 1536  
RESULT 5  
CAL4 HUMAN  
ID CAL4 HUMAN STANDARD; PRT; 1669 AA.  
AC P02462;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(IV) chain precursor.  
GN COL4A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89340433; PubMed=2701944;  
RA Soiminen R., Huctari M., Ganguly A., Prockop D.J., Tryggvason K.;  
RT "Structural organization of the gene for the alpha 1 chain of human  
RT type IV collagen.";

RL J. Biol. Chem. 264:13565-13571 (1989).  
RN [2]  
RP SEQUENCE OF 46-1257 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88083584; PubMed=3691802;  
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
RT "Complete primary structure of the alpha 1-chain of human basement  
RT membrane (type IV) collagen.";  
RL FEBS Lett. 225:188-194 (1987).  
RN [3]  
RP SEQUENCE OF 1-943 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=89029471; PubMed=3311751;  
RA Brazel D., Oberbaumer I., Dieringer H., Babel W., Glanville R.W.,  
RA Deutzmann R., Kuehn K.;  
RT "Completion of the amino acid sequence of the alpha 1 chain of human  
RT basement membrane collagen (type IV) reveals 21 non-triplet  
RT interruptions located within the collagenous domain.";  
RL Eur. J. Biochem. 168:529-536 (1987).  
RN [4]  
RP SEQUENCE OF 28-243.  
RX MEDLINE=86004708; PubMed=4043082;  
RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
RT region (7S domain) of the alpha 1 (IV) chain of human basement  
RT membrane collagen.";  
RL Eur. J. Biochem. 152:213-219 (1985).  
RN [5]  
RP SEQUENCE OF 534-1447.  
RX MEDLINE=85003629; PubMed=6434307;  
RA Babel W., Glanville R.W.;  
RT "Structure of human basement-membrane (type IV) collagen. Complete  
RT amino-acid sequence of a 914-residue-long pepsin fragment from the  
RT alpha 1(IV) chain.";  
RL Eur. J. Biochem. 143:545-556 (1984).  
RN [6]  
RP SEQUENCE OF 1256-1669 FROM N.A.  
RX MEDLINE=85207819; PubMed=2581969;  
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
RA Chung M.-C., Prockop D.J., Boyd C.D.;  
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV  
RT procollagen reveal an unusual homology of amino acid sequences in two  
RT halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687 (1985).  
RN [7]  
RP SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RA Katalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
RT procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653 (1985).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89034231; PubMed=2182844;  
RA Soininen R., Huotari M., Hostikka S.L., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
RT collagen are divergently encoded on opposite DNA strands and have an  
RT overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220 (1988).  
RN [9]  
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RT of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624 (1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M26576; AAA53098.1; JOINED.  
CC EMBL; J04217; AAA53098.1; JOINED.  
CC EMBL; M26550; AAA53098.1; JOINED.  
CC EMBL; M26540; AAA53098.1; JOINED.  
CC EMBL; M26542; AAA53098.1; JOINED.  
CC EMBL; M26543; AAA53098.1; JOINED.  
CC EMBL; M26544; AAA53098.1; JOINED.  
CC EMBL; M26545; AAA53098.1; JOINED.  
CC EMBL; M26546; AAA53098.1; JOINED.  
CC EMBL; M26547; AAA53098.1; JOINED.  
CC EMBL; M26537; AAA53098.1; JOINED.  
CC EMBL; M26538; AAA53098.1; JOINED.  
CC EMBL; M26548; AAA53098.1; JOINED.  
CC EMBL; M26549; AAA53098.1; JOINED.  
CC EMBL; M26551; AAA53098.1; JOINED.  
CC EMBL; M26552; AAA53098.1; JOINED.  
CC EMBL; M26553; AAA53098.1; JOINED.  
CC EMBL; M26554; AAA53098.1; JOINED.  
CC EMBL; M26555; AAA53098.1; JOINED.  
CC EMBL; M26556; AAA53098.1; JOINED.  
CC EMBL; M26557; AAA53098.1; JOINED.  
CC EMBL; M26539; AAA53098.1; JOINED.  
CC EMBL; M26558; AAA53098.1; JOINED.  
CC EMBL; M26559; AAA53098.1; JOINED.  
CC EMBL; M26560; AAA53098.1; JOINED.  
CC EMBL; M26561; AAA53098.1; JOINED.  
CC EMBL; M26562; AAA53098.1; JOINED.  
CC EMBL; M26536; AAA53098.1; JOINED.  
CC EMBL; M26563; AAA53098.1; JOINED.  
CC EMBL; M26541; AAA53098.1; JOINED.  
CC EMBL; M26564; AAA53098.1; JOINED.  
CC EMBL; M26565; AAA53098.1; JOINED.  
CC EMBL; M26566; AAA53098.1; JOINED.  
CC EMBL; M26567; AAA53098.1; JOINED.  
CC EMBL; M26568; AAA53098.1; JOINED.  
CC EMBL; M26569; AAA53098.1; JOINED.  
CC EMBL; M26570; AAA53098.1; JOINED.  
CC EMBL; M26571; AAA53098.1; JOINED.  
CC EMBL; M26572; AAA53098.1; JOINED.  
CC EMBL; M26573; AAA53098.1; JOINED.  
CC EMBL; M26574; AAA53098.1; JOINED.  
CC EMBL; M26575; AAA53098.1; JOINED.  
CC EMBL; Y00706; CAA58698.1; -  
CC EMBL; X05561; CAA29075.1; -  
CC EMBL; M10940; AAA52006.1; -  
CC EMBL; M11315; AAA52042.1; -  
CC PIR; S16876; CGHU4B.  
CC Genew; HGNC:2202; COL4A1.

```
DR MIW; 120130; --
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126
FT DISULFID 1460 1551 N-LINKED (GLCNAC. . .).
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1562.
FT DISULFID 1570 1665 OR 1665.
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CONFLICT 237 238 SG -> XE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 E -> Q (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9BA84 CRC64;
```

Query Match 77.3%; Score 109; DB 1; Length 1669;  
Best Local Similarity 73.1%; Pred. No. 4.4e-08;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFVTMPFLFSNVDSVFASRNDYS 27  
:::|||||:|:|||||  
DB 1495 RKFTMPFLFCNINNVCFASRNDYS 1520

```
RESULT 6
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=89112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
```

```
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Ghanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quirones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J03758; AAA37439.1; -;  
 DR EMBL; M23333; AAA51625.1; -;  
 DR EMBL; J04694; AAA50292.1; -;  
 DR EMBL; X06777; CAA29946.1; -;  
 DR EMBL; X02201; CAA26132.1; -;  
 DR EMBL; M15832; AAA37340.1; -;  
 DR EMBL; M14042; AAA37342.1; -;  
 DR EMBL; M12879; AAA37343.1; -;  
 DR EMBL; M13024; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M13025; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M13026; AAA37344.1; -;  
 DR EMBL; M13027; AAA37345.1; -;  
 DR EMBL; M13043; AAA37346.1; -;  
 DR EMBL; J04448; AAA37437.1; -;  
 DR EMBL; J03525; CGMS4B.  
 DR MGD; MGI:88454; Col4a1.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 23.  
 DR ProDom; PD000007; Clg\_helix; 6.  
 DR ProDom; PD003923; Procollagn4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Repeat; Hydroxylation; Connective tissue; Basement membrane;  
 KW Extracellular matrix; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT NONHELIICAL REGION (NCL).  
 FT OR 1441 1669  
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
 FT OR 1551 (BY SIMILARITY).  
 FT BY SIMILARITY.  
 FT OR 1662 (BY SIMILARITY).  
 FT OR 1665 (BY SIMILARITY).  
 FT BY SIMILARITY.  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 126 126  
 FT CARBOHYD 126 126  
 FT DISULFID 126 126  
 FT CONFLICT 26 26 A -> P (IN REF. 2).  
 FT CONFLICT 186 186 S -> L (IN REF. 2).  
 FT CONFLICT 319 319 Q -> S (IN REF. 2).  
 FT CONFLICT 369 369 L -> F (IN REF. 2).  
 FT CONFLICT 403 403 P -> L (IN REF. 2).  
 FT CONFLICT 481 481 Q -> H (IN REF. 2).  
 FT CONFLICT 493 493 E -> I (IN REF. 2).  
 FT CONFLICT 712 712 E -> Q (IN REF. 2).  
 FT CONFLICT 813 813 Q -> H (IN REF. 2).  
 FT CONFLICT 982 982 V -> S (IN REF. 2).  
 FT CONFLICT 1397 1397  
 FT SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;  
 Query Match 77.3%; Score 109; DB 1; Length 1669;  
 Best Local Similarity 73.1%; Pred. No. 4.4e-08;  
 Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 QRFTMPFLFNSVNDVSNFASRNDVS 27  
 Db 1495 RKFSTMPFLFCNNVNCVFNASRNDVS 1520  
 RESULT 7  
 CA24\_ASCSU STANDARD; PRT; 1763 AA.  
 ID CA24\_ASCSU  
 AC P27393;  
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
 OC Ascarididae; Ascaris.  
 OC NCBI\_TaxID=6253;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).  
 RX MEDLINE=91340768; PubMed=1714907;  
 RA Pettitt J.; Kingston I.B.;  
 RT "The complete primary structure of a nematode alpha 2(IV) collagen  
 and the partial structural organization of its gene.";  
 RL J. Biol. Chem. 266:16149-16156(1991).  
 CC !- FUNCTION: Collagen type IV is specific for basement membranes.  
 CC !- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
 CC Type IV collagen forms a mesh-like network linked through  
 CC intermolecular interactions between 7S domains and between NC1  
 CC domains.  
 CC !- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=I;  
 CC IsoId=P27393-1; Sequence=Displayed;  
 CC Name=II;  
 CC IsoId=P27393-2; Sequence=VSP\_001159;  
 CC !- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the  
 CC G-X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC !- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC !- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M67507; AA18014.1; -;  
 DR PIR; S16366; S16366.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 25.  
 DR ProDom; PD000007; Clg\_helix; 6.  
 DR ProDom; PD003923; Procollagn4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;  
 KW Alternative splicing; Glycoprotein; Signal.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 27 42 7S DOMAIN.  
 FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.  
 FT NONHELIICAL REGION (NCL).  
 FT OR 1530 1763  
 FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).  
 FT OR 1637 (BY SIMILARITY).  
 FT BY SIMILARITY.  
 FT DISULFID 1581 1599 OR 1749 (BY SIMILARITY).  
 FT OR 1749 (BY SIMILARITY).  
 FT DISULFID 1656 1752 OR 1752 (BY SIMILARITY).  
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).  
 FT DISULFID 1702 1709 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT OR 126 126 O-LINKED (XYL. . .) (GLYCOSAMINOGLYCAN)  
 FT 249 249 (IN ISOFORM II) (POTENTIAL).  
 FT VARSPLIC 230 266  
 FT GLOGPGRGPGPGPSTGAKTITGPEGAPGMKGEK ->  
 FT GLIGPAGPGRPGPSTGAKTITGPEGHSGDKGVK (in





```
Matches 17; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFSNVNDSNFASRNDYS 27
Db 1581 QRFTHMPFLFSNVNDSNFASRNEKS 1606

RESULT 9
CA14_CAEEL STANDARD; PRT; 1758 AA.
ID CA14_CAEEL
AC P17139;
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV), chain precursor.
GN EMB-9 OR CIB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peliceridae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=91141582; PubMed=1996137;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RL Nature 349:707-709(1991)."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirksen J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans."
RL Nature 368:32-38(1994)."
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RL genes are located on separate chromosomes."
RL J. Biol. Chem. 264:17574-17582(1989)."
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Eye, and Kidney;  
RX MEDLINE=94171779; PubMed=8125972;  
RA Ohashi T., Sugimoto M., Mattai M.-G., Ninomiya Y.;  
RT identification of a new collagen IV chain, alpha 6(IV), by cDNA  
RT isolation and assignment of the gene to chromosome Xq22, which is the  
RT same locus for COL4A5.";  
RL J. Biol. Chem. 269:7520-7526(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=94230418; PubMed=8175748;  
RA Zhou J., Ding M., Zhao Z., Reeder S.T.;  
RT "Complete primary structure of the sixth chain of human basement  
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)  
RT and comparison with five other type IV collagen chains.";  
RL J. Biol. Chem. 269:13193-13199(1994).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND  
RP LYS-1110.  
RX MEDLINE=96299642; PubMed=8661006;  
RA Zhang X., Zhou J., Reeder S.T., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Bird C., Grafham D., Lawlor S., Wilson S.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=93361972; PubMed=8356449;  
RA Zhou J., Nishizaki T., Smets H., Antignac C., Laurila P.,  
RA de Paeppe A., Tryggvason K., Reeder S.T.;  
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RT inherited smooth muscle tumors.";  
RL Science 261:1167-1169(1993).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM) forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing. Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequence=VSP\_001174;  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D21337; BAA04809.1; -;  
DR EMBL; U04845; AAA19569.2; -;  
DR EMBL; U47004; AAB19038.1; JOINED.  
DR EMBL; U46959; AAB19038.1; JOINED.  
DR EMBL; U46961; AAB19038.1; JOINED.  
DR EMBL; U46962; AAB19038.1; JOINED.  
DR EMBL; U46963; AAB19038.1; JOINED.  
DR EMBL; U46964; AAB19038.1; JOINED.  
DR EMBL; U46965; AAB19038.1; JOINED.  
DR EMBL; U46966; AAB19038.1; JOINED.  
DR EMBL; U46967; AAB19038.1; JOINED.  
DR EMBL; U46968; AAB19038.1; JOINED.  
DR EMBL; U46969; AAB19038.1; JOINED.  
DR EMBL; U46970; AAB19038.1; JOINED.  
DR EMBL; U46971; AAB19038.1; JOINED.  
DR EMBL; U46972; AAB19038.1; JOINED.  
DR EMBL; U46973; AAB19038.1; JOINED.  
DR EMBL; U46974; AAB19038.1; JOINED.  
DR EMBL; U46975; AAB19038.1; JOINED.  
DR EMBL; U46976; AAB19038.1; JOINED.  
DR EMBL; U46977; AAB19038.1; JOINED.  
DR EMBL; U46978; AAB19038.1; JOINED.  
DR EMBL; U46979; AAB19038.1; JOINED.  
DR EMBL; U46980; AAB19038.1; JOINED.  
DR EMBL; U46981; AAB19038.1; JOINED.  
DR EMBL; U46982; AAB19038.1; JOINED.  
DR EMBL; U46983; AAB19038.1; JOINED.  
DR EMBL; U46984; AAB19038.1; JOINED.  
DR EMBL; U46985; AAB19038.1; JOINED.  
DR EMBL; U46986; AAB19038.1; JOINED.  
DR EMBL; U46987; AAB19038.1; JOINED.  
DR EMBL; U46988; AAB19038.1; JOINED.  
DR EMBL; U46989; AAB19038.1; JOINED.  
DR EMBL; U46990; AAB19038.1; JOINED.  
DR EMBL; U46991; AAB19038.1; JOINED.  
DR EMBL; U46992; AAB19038.1; JOINED.  
DR EMBL; U46993; AAB19038.1; JOINED.  
DR EMBL; U46994; AAB19038.1; JOINED.  
DR EMBL; U46995; AAB19038.1; JOINED.  
DR EMBL; U46996; AAB19038.1; JOINED.  
DR EMBL; U46997; AAB19038.1; JOINED.  
DR EMBL; U46998; AAB19038.1; JOINED.  
DR EMBL; U46999; AAB19038.1; JOINED.  
DR EMBL; U47000; AAB19038.1; JOINED.  
DR EMBL; U47001; AAB19038.1; JOINED.  
DR EMBL; U47002; AAB19038.1; JOINED.  
DR EMBL; U47003; AAB19038.1; JOINED.  
DR EMBL; U47004; AAB19039.1; -;  
DR EMBL; U46960; AAB19039.1; JOINED.  
DR EMBL; U46961; AAB19039.1; JOINED.  
DR EMBL; U46962; AAB19039.1; JOINED.  
DR EMBL; U46963; AAB19039.1; JOINED.  
DR EMBL; U46964; AAB19039.1; JOINED.  
DR EMBL; U46965; AAB19039.1; JOINED.  
DR EMBL; U46966; AAB19039.1; JOINED.  
DR EMBL; U46967; AAB19039.1; JOINED.  
DR EMBL; U46968; AAB19039.1; JOINED.  
DR EMBL; U46969; AAB19039.1; JOINED.  
DR EMBL; U46970; AAB19039.1; JOINED.  
DR EMBL; U46971; AAB19039.1; JOINED.  
DR EMBL; U46972; AAB19039.1; JOINED.  
DR EMBL; U46973; AAB19039.1; JOINED.  
DR EMBL; U46974; AAB19039.1; JOINED.  
DR EMBL; U46975; AAB19039.1; JOINED.  
DR EMBL; U46976; AAB19039.1; JOINED.  
DR EMBL; U46977; AAB19039.1; JOINED.  
DR EMBL; U46978; AAB19039.1; JOINED.  
DR EMBL; U46979; AAB19039.1; JOINED.  
DR EMBL; U46980; AAB19039.1; JOINED.  
DR EMBL; U46981; AAB19039.1; JOINED.  
DR EMBL; U46982; AAB19039.1; JOINED.  
DR EMBL; U46983; AAB19039.1; JOINED.

DR EMBL; U45984; AAB19039.1; JOINED.  
 DR EMBL; U45985; AAB19039.1; JOINED.  
 DR EMBL; U45986; AAB19039.1; JOINED.  
 DR EMBL; U45987; AAB19039.1; JOINED.  
 DR EMBL; U45988; AAB19039.1; JOINED.  
 DR EMBL; U45989; AAB19039.1; JOINED.  
 DR EMBL; U45990; AAB19039.1; JOINED.  
 DR EMBL; U45991; AAB19039.1; JOINED.  
 DR EMBL; U45992; AAB19039.1; JOINED.  
 DR EMBL; U45993; AAB19039.1; JOINED.  
 DR EMBL; U45994; AAB19039.1; JOINED.  
 DR EMBL; U45995; AAB19039.1; JOINED.  
 DR EMBL; U45996; AAB19039.1; JOINED.  
 DR EMBL; U45997; AAB19039.1; JOINED.  
 DR EMBL; U45998; AAB19039.1; JOINED.  
 DR EMBL; U45999; AAB19039.1; JOINED.  
 DR EMBL; U47000; AAB19039.1; JOINED.  
 DR EMBL; U47001; AAB19039.1; JOINED.  
 DR EMBL; U47002; AAB19039.1; JOINED.  
 DR EMBL; U47003; AAB19039.1; JOINED.  
 DR EMBL; AL034369; CAA22265.1; -.  
 DR EMBL; AL109943; CAB99263.1; -.  
 DR EMBL; AL136080; CAB96748.1; -.  
 DR EMBL; AL031177; CAA20120.1; -.  
 DR EMBL; L22763; AAA16338.1; -.  
 DR PIR; A54122; CGHU6B.  
 DR Genew; HGNC:2208; COL4A6.  
 DR MIM; 303631; -.  
 DR GO; GO:0005587; C:collagen type IV; NAS.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.  
 DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.  
 DR InterPro; IPR008161; C1g helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 23.  
 DR ProDom; PDC000007; C1g\_helix; 4.  
 DR ProDom; PDC03923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Basement membrane;  
 KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;  
 KW Alternative splicing; Polymorphism.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.  
 FT DOMAIN 23 46 7S DOMAIN.  
 Query Match 60.3%; Score 85; DB 1; Length 1691;  
 Best Local Similarity 56.0%; Pred. NO. 0.00017;  
 Matches 14; Conservative 7; Mismatches 4; Indels 0; Gaps 0;  
 OY 3 RPTMPLFSLNVDNVSFASNDYS 27  
 DB 1518 RPTMPLFYCINEVCHVARENDXS 1542  
 RESULT 11  
 CR224 MOUSE  
 ID CA24\_MOUSE STANDARD; PRT; 1707 AA.  
 AC P08122; Q61375;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 GN COL4A2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=8197933; PubMed=2703491;  
 RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,  
 RA Pihlajaniemi T., Kurkinen M.;  
 "The complete primary structure of mouse alpha 2(IV) collagen.  
 Alignment with mouse alpha 1(IV) collagen.";  
 J. Biol. Chem. 264:6318-6324 (1989)."  
 [2]  
 RP MEDLINE=89066738; PubMed=3198626;  
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogel G.;  
 "Head-to-head arrangement of murine type IV collagen genes.";  
 J. Biol. Chem. 263:19274-19277 (1988)."  
 [3]  
 RP MEDLINE=970-1480 FROM N.A.  
 RA MEDLINE=86220192; PubMed=3011432;  
 RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,  
 RA Deutzmann R., Timpl R., Kuehn K.;  
 "Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain.";  
 Eur. J. Biochem. 157:49-56 (1986)."  
 [4]  
 RP MEDLINE=1480-1707 FROM N.A.  
 RA MEDLINE=87054581; PubMed=3780963;  
 RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;  
 "cDNA and protein sequence of the NCI domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV).";  
 FEBS Lett. 208:203-207 (1986)."  
 [5]  
 RP MEDLINE=1481-1707 FROM N.A.  
 RA MEDLINE=87250460; PubMed=3597383;  
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
 RA Saus J., Pihlajaniemi T.;  
 "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";  
 J. Biol. Chem. 262:8496-8499 (1987)."  
 [6]  
 RP MEDLINE=1041-1489 FROM N.A.  
 RA MEDLINE=87005245; PubMed=3758345;  
 RA Vogel G., Horn E., Carter J., Kaytes P.S.;  
 "Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";  
 FEBS Lett. 206:29-32 (1986)."  
 [7]  
 RP MEDLINE=964-1003; 1005-1085 AND 1087-1109 FROM N.A.  
 RA MEDLINE=85296379; PubMed=3839908;  
 RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;  
 "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";  
 Nature 317:177-179 (1985)."  
 [8]  
 RP MEDLINE=1-60 FROM N.A.  
 RA MEDLINE=89071759; PubMed=3200851;  
 RA Burdello P.D., Martin G.R., Yamada Y.;  
 "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";  
 Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988)."  
 CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) - alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M23334; AAA51626.1; JOINED.  
 DR EMBL; M23333; AAA51626.1; JOINED.  
 DR EMBL; J04695; AAA50293.1;  
 DR EMBL; J04448; AAA37438.1;  
 DR EMBL; J04647; CAA28308.1;  
 DR EMBL; M15833; AAA37341.1;  
 DR EMBL; X04410; CAA27998.1;  
 DR EMBL; X02896; CAA26655.1;  
 DR EMBL; X02897; CAA51614.1;  
 DR EMBL; X02898; CAA26657.1;  
 DR EMBL; X02899; CAA26658.1;  
 DR PIR; A33526; A33526.  
 DR MGI; 88455; Col4a2.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.  
 DR ProDom; PD000007; C1g\_helix; 7.  
 DR ProDom; PD003923; Procollagen4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR GlycoProtein; Basement membrane; Collagen; Signal.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 FT SIGNAL 1 25 AVINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 26 183 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 184 1707 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1480 1707 NON-HELICAL REGION (NC1).  
 FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).  
 FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).  
 FT DISULFID 1547 1550 BY SIMILARITY.  
 FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).  
 FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).  
 FT DISULFID 1653 1660 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1051 1051 P -> R (IN REF. 6).  
 FT CONFLICT 1097 1097 S -> G (IN REF. 7).  
 FT CONFLICT 1171 1171 G -> S (IN REF. 6).  
 FT CONFLICT 1179 1179 P -> R (IN REF. 6).  
 FT CONFLICT 1241 1241 Q -> E (IN REF. 6).  
 FT CONFLICT 1328 1328 P -> A (IN REF. 6).  
 FT CONFLICT 1573 1573 V -> L (IN REF. 4).  
 FT CONFLICT 1623 1623 Y -> H (IN REF. 4).  
 SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;  
 Query Match 58.2%; Score 82; DB 1; Length 1707;  
 Best Local Similarity 64.0%; Pred. No. 0.00047;  
 Matches 16; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 3 RFTMTPLFNVNDVSNFASRNDYS 27  
 DB 1535 RFTMTPLFNVNDVSNFASRNDYS 1559  
 RESULT 12  
 CA24 HUMAN  
 ID CA24 HUMAN STANDARD; PRT; 1712 AA.  
 AC P08572;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 GN COL4A2.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89066769; PubMed=3198637;  
 RA Hostikka S.L., Tryggvason K.;  
 RT "The complete primary structure of the alpha 2 chain of human type IV  
 RL collagen and comparison with the alpha 1(IV) chain.";  
 RL J. Biol. Chem. 263:19488-19493(1988).  
 RN [2]  
 RP SEQUENCE OF 1-1042 FROM N.A.  
 RX MEDLINE=8815198; PubMed=3345760;  
 RA Hostikka S.L., Pollner R., Oberbauer I., Kuehn K.;  
 RT "Human basement membrane collagen (type IV). The amino acid sequence  
 RL of the alpha 2(IV) chain and its comparison with the alpha 1(IV)  
 RL chain reveals deletions in the alpha 1(IV) chain.";  
 RL Eur. J. Biochem. 172:35-42(1988).  
 RN [3]  
 RP SEQUENCE OF 1254-1712 FROM N.A.  
 RX MEDLINE=87219158; PubMed=3582677;  
 RA Hostikka S.L., Kurkinen M., Tryggvason K.;  
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2  
 RL chain cDNA reveals extensive homology with the NC-1 domain of alpha 1  
 RL (IV) but not with the collagenous domain or 3'-untranslated region.";  
 RL FEBS Lett. 216:281-286(1987).  
 RN [4]  
 RP SEQUENCE OF 1451-1485 FROM N.A.  
 RX MEDLINE=87092438; PubMed=3025878;  
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;  
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and  
 RL alpha 2 (IV) chains map to the distal long arm of chromosome 13.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).  
 RN [5]  
 RP SEQUENCE OF 1486-1712 FROM N.A.  
 RX MEDLINE=87250571; PubMed=2439508;  
 RA Myer J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;  
 RT "Duplication of type IV collagen COOH-terminal repeats and species-  
 RL specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";  
 RL J. Biol. Chem. 262:9231-9238(1987).  
 RN [6]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89034231; PubMed=3182844;  
 RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
 RL collagen are divergently encoded on opposite DNA strands and have an  
 RL overlapping promoter region.";  
 RL J. Biol. Chem. 263:17217-17220(1988).  
 RN [7]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89030632; PubMed=2846280;  
 RA Poeschl E., Pollner R., Kuehn K.;  
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human  
 RL basement membrane collagen type IV are arranged head-to-head and  
 RL separated by a bidirectional promoter of unique structure.";  
 RL EMBO J. 7:2687-2695(1988).  
 RN [8]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=93305049; PubMed=8317999;  
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;  
 RT "Identification of a novel sequence element in the common promoter  
 RL region of human collagen type IV genes, involved in the regulation of  
 RL divergent transcription.";  
 RL Biochem. J. 292:667-695(1993).  
 RN [9]  
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.  
 RX MEDLINE=89005112; PubMed=2844531;  
 RA Siebold B., Deutzmann R., Kuehn K.;  
 RT "The arrangement of intra- and intermolecular disulfide bonds in the

RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RL of basement-membrane type IV collagen.;;  
CC Eur. J. Biochem. 176:617-624(1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05562; CAA29076.1; -;  
CC EMBL; X05610; CAA29098.1; -;  
CC EMBL; J02760; AAA58422.1; -;  
CC EMBL; M36963; AAA53099.1; -;  
CC EMBL; X12784; CAA31275.1; -;  
CC EMBL; J04217; AAA53097.1; -;  
CC PIR; A32024; CGHU2B  
CC Genew; HGNC:2203; COL4A2.  
CC MIM; 120090; -;  
CC DR GO; GO:0005587; C:collagen type IV; TAS.  
CC DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.  
CC DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.  
CC DR InterPro; IPR008161; C1q helix.  
CC DR InterPro; IPR008160; Collagen.  
CC DR InterPro; IPR001442; Procollagn4\_C.  
CC DR Pfam; PF01413; C4; 2.  
CC DR ProDom; PD000007; C1q helix; 7.  
CC DR ProDom; PD003923; ProcollagnC4; 1.  
CC DR SMART; SM00111; C4; 2.  
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Signal.  
FT SIGNAL 1 25  
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.  
FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.  
FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).  
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).  
FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).  
FT DISULFID 1549 1555 BY SIMILARITY.  
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).  
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).  
FT DISULFID 1658 1665 BY SIMILARITY.  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .).  
FT CONFLICT 471 471 R -> P (IN REF. 2).  
FT CONFLICT 683 683 A -> G (IN REF. 2).  
FT CONFLICT 1575 1575 M -> I (IN REF. 5).  
FT CONFLICT 1663 1663 G -> H (IN REF. 9).  
FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
FT CONFLICT 1712 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
FT SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
Query Match 58.2%; Score 82; DB 1; Length 1712;  
Best Local Similarity 64.0%; Pred. No. 0.00047;

Matches 16; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 3 RFTMPPLFSNVNDVSFASNDYS 27  
DB 1540 RFTMPPLFGVGVYASRNDKS 1564  
RESULT 13  
CA44. RABIT  
ID CA44 RABIT STANDARD; PRT; 623 AA.  
AC P55787;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Collagen alpha 4(IV) chain (Fragment).  
GN COL4A4.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Corneal endothelium;  
RX MEDLINE=93054733; PubMed=1429714;  
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;  
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
RT alpha 4 chain of basement membrane collagen type IV and assignment of  
RT the gene to the distal long arm of human chromosome 2.";  
RL J. Biol. Chem. 267:23753-23758(1992).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; J01477; -; NOT\_ANNOTATED\_CDS.  
CC FIR; A45137; A45137.  
CC InterPro; IPR008160; Collagen.  
CC InterPro; IPR001442; Procollagn4\_C.  
CC Pfam; PF01413; C4; 2.  
CC Pfam; PF01391; Collagen; 5.  
CC ProDom; PD003923; ProcollagnC4; 1.  
CC SMART; SM00111; C4; 2.  
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
FT NON\_TER 1 1  
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.  
FT DOMAIN 393 623 NONHELICAL REGION (NC1).  
FT DISULFID 413 502 OR 499 (BY SIMILARITY).  
FT DISULFID 446 499 OR 502 (BY SIMILARITY).





3

DR EMBL; M28334; AAA28422.1; -;  
DR EMBL; V00200; CAA23486.2; -;  
DR PIR; A31893; A31893.  
DR FlyBase; FBgn0000299; Cg25C.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 25.  
DR ProDom; PD000007; Clg\_helix; 9.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 23  
FT PROPEP 24 ?  
FT CHAIN ? 1775  
FT DOMAIN ? 1544  
FT DOMAIN 1545 1775  
FT DISULFID 1569 1655  
FT DISULFID 1599 1652  
FT DISULFID 1611 1617  
FT DISULFID 1674 1770  
FT DISULFID 1708 1767  
FT DISULFID 1720 1727  
FT CARBOHYD 72 72  
FT CONFLICT 948 948  
FT CONFLICT 997 997  
FT CONFLICT 1357 1357  
FT CONFLICT 1360 1360  
FT CONFLICT 1373 1373  
FT CONFLICT 1496 1496  
FT CONFLICT 1507 1511  
FT CONFLICT 1529 1529  
FT CONFLICT 1733 1733  
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AE23149525CD CRC64;  
AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
COLLAGEN ALPHA 1(IV) CHAIN.  
TRIPLE-HELICAL REGION.  
NONHELICAL REGION (NC1).  
OR 1652 (BY SIMILARITY).  
OR 1655 (BY SIMILARITY).  
BY SIMILARITY.  
OR 1767 (BY SIMILARITY).  
OR 1770 (BY SIMILARITY).  
BY SIMILARITY.  
N-LINKED (GLCNAC. . .) (PROBABLE).  
L -> S (IN REF. 6).  
S -> T (IN REF. 6).  
Q -> K (IN REF. 5).  
T -> I (IN REF. 5).  
L -> R (IN REF. 5).  
ETGNV -> RAGOR (IN REF. 5).  
E -> K (IN REF. 5).  
M -> I (IN REF. 5).

Query Match 48.2%; Score 68; DB 1; Length 1775;  
Best Local Similarity 56.5%; Pred.No. 0.06;  
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 RFTTPEPLFSNVNDVNSFNRND 25  
||:|:|  
Db 1602 RFSTLPVLSGQNNVCNYSRND 1624

Search completed: April 5, 2004, 06:59:40  
Job time : 4.39952 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 Seconds  
(without alignments)  
525.440 Million cell updates/sec

Title: US-10-032-221B-40  
Perfect score: 141  
Sequence: 1 KQFTTWPFLFSNVNDVSNFASNDYS 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_nhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rident:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	89.4	212	6 Q28512	Q28512 macaca mula
2	126	89.4	245	4 Q9NYC4	Q9NYC4 homo sapien
3	125	88.7	203	6 Q29032	Q29032 sus scrofa
4	125	88.7	203	6 Q28682	Q28682 oryctolagus
5	125	88.7	212	6 Q28587	Q28587 ovis aries
6	120	85.1	161	11 Q61430	Q61430 mus musculus
7	120	85.1	210	6 Q28273	Q28273 canis faml
8	120	85.1	246	11 Q61435	Q61435 mus musculus
9	120	85.1	1669	11 Q9QZS0	Q9QZS0 mus musculus
10	116	82.3	230	11 Q63122	Q63122 rattus norv
11	110	78.0	179	11 P70165	P70165 mus musculus
12	110	78.0	253	11 Q61436	Q61436 mus musculus
13	110	78.0	585	11 Q80V57	Q80V57 mus musculus
14	110	78.0	799	11 Q8BNS7	Q8BNS7 mus musculus
15	110	78.0	886	4 Q9NUB7	Q9NUB7 homo sapien
16	110	78.0	1684	6 Q8HYC1	Q8HYC1 canis faml

17	110	78.0	1688	6 Q86622	Q86622 canis faml
18	110	78.0	1691	11 Q9ESQ2	Q9ESQ2 mus musculu
19	109	77.3	225	6 Q28271	Q28271 canis faml
20	109	77.3	226	11 Q99LQ8	Q99LQ8 mus musculu
21	109	77.3	229	4 Q8NF88	Q8NF88 homo sapien
22	109	77.3	229	4 Q9NYC5	Q9NYC5 homo sapien
23	109	77.3	979	13 Q919K3	Q919K3 gallus gall
24	109	77.3	1075	4 Q86X41	Q86X41 homo sapien
25	109	77.3	1621	4 Q9H4R9	Q9H4R9 homo sapien
26	100	70.9	1747	5 Q26640	Q26640 strongyloce
27	100	70.9	1752	5 Q07265	Q07265 strongyloce
28	88	62.4	1802	5 Q17163	Q17163 brugia mala
29	85	60.3	205	6 Q28274	Q28274 canis faml
30	85	60.3	546	11 Q99K97	Q99K97 mus musculu
31	85	60.3	1600	4 Q9UEH6	Q9UEH6 homo sapien
32	85	60.3	1691	11 Q9ESQ1	Q9ESQ1 mus musculu
33	82	58.2	202	6 Q28272	Q28272 canis faml
34	82	58.2	358	11 Q91V13	Q91V13 mus musculu
35	82	58.2	673	4 Q14052	Q14052 homo sapien
36	79	56.0	1723	5 Q9GQB1	Q9GQB1 hydra atten
37	69	48.9	1761	5 Q18407	Q18407 drosophila
38	69	48.9	1940	5 Q3VMV5	Q3VMV5 drosophila
39	68	48.2	312	11 Q64457	Q64457 mus musculu
40	68	48.2	1682	11 Q9QZR9	Q9QZR9 mus musculu
41	68	48.2	1779	5 Q9VMV4	Q9VMV4 drosophila
42	67	47.5	208	6 Q29468	Q29468 canis faml
43	67	47.5	1024	5 Q8T7S4	Q8T7S4 anopheles g
44	64	45.4	713	5 Q5GV24	Q5GV24 sarcophaga
45	59	41.8	1004	10 Q9LGN1	Q9LGN1 cryza eativ

#### ALIGNMENTS

#### RESULT 1

Q28512	PRELIMINARY;	PRT;	212 AA.
ID	Q28512		
AC	Q28512;		
DT	01-NOV-1996 (TRENBLrel. 01, Created)		
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Alpha-3 type IV collagen (Fragment).		
GN	COL4A3.		
OS	Macaca mulatta (Rhesus macaque).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;		
OC	Cercopitheidae; Macaca.		
OX	NCBI_TaxID=9544;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Kidney cortex;		
RA	Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,		
RA	Mason P.J., Pusey C.D.;		
RT	"Properties and sequences of the Goodpasture antigen of different		
RT	mammals."		
RL	Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; I47280; AA91861.1; -		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	InterPro; IPR001442; Procollagn4_C.		
DR	InterPro; IPR000504; RNA_rec_mot.		
DR	Fram; PF01413; C4; 2.		
DR	Prodom; PD003923; ProcollagnC4; 1.		
DR	SMART; SM00111; C4; 2.		
DR	PROSITE; PS00030; RRM_RNP_1; 1.		
KW	Collagen.		
FT	NON_TER	1	
FT	NON_TER	212	
SQ	SEQUENCE	212 AA;	4BC574A64E357E64 CRC64;

Query Match 89.4%; Score 136; DB 6; Length 212;  
Best Local Similarity 92.3%; Pred. No. 7.4e-11;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27  
|||||  
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

## RESULT 2

Q29YC4 PRELIMINARY; PRT; 245 AA.  
AC Q29YC4; (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 25, Last annotation update)  
DE Tumstatin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

SEQUENCE FROM N.A.  
RP Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.;  
RT "Distinct anti-tumor properties of a type IV collagen domain derived  
from basement membrane."  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL; AF258351; AAF72632.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
FT NON\_TER 1  
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 89.4%; Score 126; DB 4; Length 245;  
Best Local Similarity 92.3%; Pred. No. 8.7e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27  
|||||  
DB 70 QRTTTPFLFCNVNDVCFASRNDYS 95

## RESULT 3

Q29032 PRELIMINARY; PRT; 203 AA.  
AC Q29032; (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
SEQUENCE FROM N.A.  
RP Tissue=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
mammals."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L47284; AAA91882.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1  
FT NON\_TER 203  
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 88.7%; Score 125; DB 6; Length 203;  
Best Local Similarity 88.5%; Pred. No. 1e-10;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27  
|||||  
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

## RESULT 4

Q28682 PRELIMINARY; PRT; 203 AA.  
AC Q28682; (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
SEQUENCE FROM N.A.  
RP Tissue=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
mammals."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L47283; AAA91893.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1  
FT NON\_TER 203  
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 88.7%; Score 125; DB 6; Length 203;  
Best Local Similarity 88.5%; Pred. No. 1e-10;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDVSNFASRNDYS 27  
|||||  
DB 37 QRTTTPFLFCNVNDVCFASRNDYS 62

## RESULT 5

Q28567 PRELIMINARY; PRT; 212 AA.  
AC Q28567; (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).

```

GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAC1904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 88.7%; Score 125; DB 6; Length 212;
Best Local Similarity 88.5%; Pred. NO. 1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
Db |||||:|||||
37 QRFTTTPFLFCNINNVCFASRNDYS 62

RESULT 6
Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the NCI domains to the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236C5 CRC64;

Query Match 85.1%; Score 120; DB 11; Length 161;
Best Local Similarity 84.6%; Pred. NO. 4.4e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
Db |||||:|||||
4 QRFTTTPFLFCNINNVCFASRNDYS 29

RESULT 7
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RL MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng X., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot-.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 85.1%; Score 120; DB 6; Length 210;
Best Local Similarity 84.6%; Pred. NO. 5.8e-10;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27
Db |||||:|||||
47 QRFTTTPFLFCNINNVCFASRNDYS 72

RESULT 8
Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c;
RA Miner J.H., Sanes J.R.;
RL MEDLINE=95050557; PubMed=7962065;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```

RT laminae: Sequence, distribution, association with laminins, and  
RT developmental switches.";  
RL J. Cell Biol. 127:879-891(1994).  
RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=Balb/c;

RA Miner J.H.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z35166; CAA84529.1; -.

DR PIR: I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO: GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR001442; Procollagn4 C.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.

FT NON\_TER

SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 85.1%; Score 120; DB 11; Length 246;

Best Local Similarity 84.6%; Pred. No. 6.9e-10;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFSNVNDYNSFASRNDYS 27

Db 71 QRTTTPFLFCNNVNCVFASRNDYS 96

RESULT 9

ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.

AC Q9QZS0;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha 3 collagen IV.

GN COL4A3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=20005934; PubMed=10534397;

RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,

RA Elder P.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;

RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a

RT mouse model of alport syndrome.";

RL Genomics 61:113-124(1999).

DR EMBL; AF169387; AAD50449.1; -.

DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.

DR GO: GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR008161; Clg\_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagn4 C.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD000007; Clg\_helix; 6.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.

KW Collagen.

SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 85.1%; Score 120; DB 11; Length 1669;

Best Local Similarity 84.6%; Pred. No. 5.4e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFSNVNDYNSFASRNDYS 27

Db 1494 QRTTTPFLFCNNVNCVFASRNDYS 1519

RESULT 10

Q63122

ID Q63122 PRELIMINARY; PRT; 230 AA.

AC Q63122;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha-3 type IV collagen (Fragment).

GN COL4A3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

EX MEDLINE=98210005; PubMed=9550634;

RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;

RT "Sequence analysis of the 'Goodpasture antigen' of mammals.";

RL Nephrol. Dial. Transplant. 13:602-607(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

RA Turner N.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; L47281; AAB72238.2; -.

DR GO: GO:0005581; C:collagen; IEA.

DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO: GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4 C.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SMO0111; C4; 2.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.

KW Collagen.

FT NON\_TER

FT NON\_TER

SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 82.3%; Score 116; DB 11; Length 230;

Best Local Similarity 84.6%; Pred. No. 2.6e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTTPFLFSNVNDYNSFASRNDYS 27

Db 55 QRTTTPFLFCNNVNCVFASRNDYS 80

RESULT 11

P70165

ID P70165 PRELIMINARY; PRT; 179 AA.

AC P70165;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Collagen type IV alpha5 chain (Fragment).

GN COL4A5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129;

RA Oberbaumer I.;

RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse

RT via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and

```

RT alpha5(IV) in differentiated teratocarcinoma cells."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82218; CAA57698.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2. Procollagn4_C.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR NON TER 1
FT NON TER 179
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 78.0%; Score 110; DB 11; Length 179;
Best Local Similarity 73.1%; Pred. No. 1.6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVNFASRNDYS 27
Db :|||||:|:|:|||||
32 RRFSTMPFMCNINNVCFASRNDYS 57

RESULT 12
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Muscle;
RX MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
RT laminae: sequence, distribution, association with laminins, and
RT developmental switches."
RL J. Cell Biol. 127:879-891(1994).
DR EMBL; Z35168; CAA84531.1; -.
DR PIR; I48304; I48304.
DR MGD; MGI:98456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 253 AA; 27626 MW; 33DAA199CA59FA91 CRC64;

Query Match 78.0%; Score 110; DB 11; Length 253;
Best Local Similarity 73.1%; Pred. No. 2.3e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVNFASRNDYS 27
Db :|||||:|:|:|||||
79 RRFSTMPFMCNINNVCFASRNDYS 104

RESULT 13
Q80V57 PRELIMINARY; PRT; 585 AA.
AC Q80V57;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Col4a5 protein.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jodan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci R., Prange C.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Sutterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043317; AAH43317.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
SQ SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 78.0%; Score 110; DB 11; Length 585;
Best Local Similarity 73.1%; Pred. No. 5.5e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVNFASRNDYS 27
Db :|||||:|:|:|||||
411 RRFSTMPFMCNINNVCFASRNDYS 436

RESULT 14
Q8ENS7 PRELIMINARY; PRT; 799 AA.
AC Q8ENS7;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Procollagen (fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```



```

60,770 full-length cDNAs.";
RL Nature 420:583-573(2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGD; MGI:88456; Col4a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMO0111; C4; 2.
FT NON_TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 78.0%; Score 110; DB 11; Length 799;
Best Local Similarity 73.1%; Pred. No. 7.7e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPLFSNVNDVNSFASRNDYS 27
DB 625 RRFSTMPFMCNINNVCFASRNDYS 650

RESULT 15
Q9NUE7 PRELIMINARY; PRT; 886 AA.
AC Q9NUE7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SMO0111; C4; 2.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 78.0%; Score 110; DB 4; Length 886;
Best Local Similarity 73.1%; Pred. No. 8.6e-08;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPLFSNVNDVNSFASRNDYS 27
DB 712 RRFSTMPFMCNINNVCFASRNDYS 737

Search completed: April 5, 2004, 07:03:58
Job time : 16.2131 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221B-40

Perfect score: 141

Sequence: 1 KQFTTWPFLFSNVNDVSNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	100.0	27	6 ADA20239	Ada20239 T8-3 pept
2	133	94.3	27	6 ADA20241	Ada20241 P2 peptid
3	131	92.9	27	6 ADA20238	Ada20238 T8 peptid
4	126	89.4	79	5 AAU75600	AAU75600 Human typ
5	126	89.4	79	6 ADA20264	ADA20264 Human tum
6	126	89.4	88	5 AAU75608	AAU75608 Human typ
7	126	89.4	88	5 AAU75607	AAU75607 Human typ
8	126	89.4	88	6 ADA20271	Ada20271 Human tum
9	126	89.4	88	6 ADA20272	Ada20272 Human tum
10	126	89.4	124	5 AAU75594	AAU75594 Human typ
11	126	89.4	124	6 ADA20258	Ada20258 Human tum
12	126	89.4	132	5 AAU75597	AAU75597 Human typ
13	126	89.4	132	6 ADA20261	Ada20261 Human tum
14	126	89.4	191	5 AAU75596	AAU75596 Human typ
15	126	89.4	191	6 ADA20260	Ada20260 Human tum
16	126	89.4	211	3 AAU95918	AAU95918 Human Goo
17	126	89.4	211	5 ABG79208	ABG79208 Human GP
18	126	89.4	218	2 AAU79164	AAU79164 Partial s
19	126	89.4	218	2 AAU44172	AAU44172 Human typ
20	126	89.4	218	3 AAU56784	AAU56784 Human alp
21	126	89.4	218	3 AAU09484	AAU09484 Human alp
22	126	89.4	232	7 ADC17697	ADC17697 Human typ
23	126	89.4	244	5 ABG79218	ABG79218 Human typ
24	126	89.4	244	5 ABG79219	ABG79219 Human Goo
25	126	89.4	244	5 ABG79217	ABG79217 Human typ

26	126	89.4	244	5 AAU75595	AAU75595 Human typ
27	126	89.4	244	6 ADA20225	Ada20225 Human typ
28	126	89.4	245	3 AAU67942	AAU67942 Human typ
29	126	89.4	245	5 AAU75589	AAU75589 Human typ
30	126	89.4	254	5 AAU75598	AAU75598 Human typ
31	126	89.4	268	2 AAU31993	AAU31993 Type IV c
32	126	89.4	268	3 AAU97555	AAU97555 Human alp
33	126	89.4	1670	7 ADD47063	ADD47063 Human pro
34	125	88.7	471	2 AAU79163	AAU79163 Partial s
35	125	88.7	471	2 AAU44171	AAU44171 Bovine ty
36	125	88.7	471	3 AAU56783	AAU56783 Bovine al
37	125	88.7	471	4 AAU09483	AAU09483 Bovine al
38	116	82.3	230	7 ADD47061	ADD47061 Rat Prote
39	110	78.0	229	7 ADC17699	ADC17699 Human typ
40	110	78.0	264	2 AAU31995	AAU31995 Type IV c
41	110	78.0	264	3 AAU97557	AAU97557 Human alp
42	110	78.0	309	3 ABU54044	ABU54044 Human pan
43	110	78.0	772	2 AAR23873	AAR23873 Human alp
44	110	78.0	772	2 AAU09643	AAU09643 Human typ
45	110	78.0	1685	4 ABG04839	ABG04839 Novel hum

## ALIGNMENTS

RESULT 1

ADA20239

ID ADA20239 standard; peptide; 27 AA.

XX ADA20239;

DT 20-NOV-2003 (first entry)

XX T8-3 peptide related to human type IV collagen alpha and angiogenesis.  
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; T8-3 peptide; tumstatin; human;  
 KW type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"

WT0003059257-A2.

24-JUL-2003.

20-DEC-2002; 2002WO-US040938.

21-DEC-2001; 2001US-00032221.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

Kalluri R;

WPI, 2003-587256/55.

New peptide, useful for preparing a composition for inhibiting tumor

growth, angiogenic activity or protein synthesis in a mammalian tissue.

Claim 63; Page 45; 240pp; English.

This invention relates to novel isolated proteins and their fragments

XX

PI Kalluri R;

iii

DR WPI; 2003-587256/55.

PT New peptide, useful for preparing a composition for inhibiting tumor

PS growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 62; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumor growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular

CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumor

CC growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present

CC sequence is the amino acid sequence of the 18 peptide of the invention,

CC derived from the amino acid sequence of tumstatin, which in turn was

CC derived from the amino acid sequence of human type IV collagen alpha 3

CC chain.

XX Sequence 27 AA;

SQ

Query Match 92.9%; Score 131; DB 6; Length 27;

Best Local Similarity 92.6%; Pred. No. 3.3e-14;

Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KQRTTMTPLFSNVNDVSNFASRNDYS 27

DB 1 KQRTTMTPLFCNVNDVSNFASRNDYS 27

RESULT 4

AAU75600

ID AAU75600 standard; protein; 79 AA.

XX AAU75600;

AC AAU75600;

XX

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tum-5.

DE

DE Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;

KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX

OS Homo sapiens.

XX

XX WO200151523-A2.

PN

XX 19-JUL-2001.

PD

XX 08-JAN-2001; 2001WO-US000565.

PF

XX 07-JAN-2000; 2000US-00479118.

PR

XX 04-APR-2000; 2000US-00543371.

PR

XX 21-JUL-2000; 2000US-00625191.

PR

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA

XX Kalluri R;

PI

XX WPI; 2002-188037/24.

DR

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

PT

PT treating disorders involving angiogenesis.

XX Example 40; Page; 205pp; English.

PS

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause

CC apoptosis of endothelial cells. Also described are the following: (1) use

CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for

CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by

CC promoting or inducing endothelial cell apoptosis in a tissue, where the

CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of

CC an antibody or peptide that specifically binds the alpha1, alpha2, or

CC alpha3, alpha4, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the

CC preparation of a medicament for inhibiting angiogenesis or cell

CC proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in the preparation

CC of a medicament for treating a proliferative disease in a vertebrate,

CC where the disease is characterised by angiogenesis that is mediated by

CC receptors to Arresten, Canstatin or Tumstatin and where the receptors

CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one

CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in

CC the presence of a medicament for promoting angiogenesis in a tissue; and

CC (5) use of integrins in the preparation of a medicament for promoting or

CC inducing angiogenesis or cell proliferation in a tissue. The fragments

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

CC or allelic variants are useful in the preparation of a medicament for

CC treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits; or by promoting or

CC inducing endothelial cell apoptosis in a tissue, where the endothelial

CC cell apoptosis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits. The medicament is useful

CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human

CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues

CC 54-132 of Tumstatin. Note: The present sequence is not shown in the

CC specification but is derived from the wild type human Tumstatin sequence

CC given in figure 18A (see AAU75589)

XX

XX Sequence 79 AA;

SQ

Query Match 89.4%; Score 126; DB 5; Length 79;

Best Local Similarity 92.3%; Pred. No. 8.3e-13;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRTTMTPLFSNVNDVSNFASRNDYS 27

DB 17 QRTTMTPLFCNVNDVSNFASRNDYS 42

RESULT 5

ADA20264

ID ADA20264 standard; protein; 79 AA.

XX ADA20264;

AC ADA20264;

XX

XX 20-NOV-2003 (first entry)

DT

XX Human tumstatin deletion protein tum-5 amino acid sequence.

DE

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

KW metastasis; basement membrane organisation; type IV collagen network;

KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX

XX Homo sapiens.

XX WO2003059257-A2.  
 XX 24-JUL-2003.  
 XX 20-DEC-2002; 2002WO-US040938.  
 XX 21-DEC-2001; 2001US-00032221.  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX Kalluri R;  
 XX WPI; 2003-587256/55.  
 XX N-PSDB; ADA20224.  
 XX New peptide, useful for preparing a composition for inhibiting tumor  
 XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX Claim 94; SEQ ID NO 26; 240pp; English.  
 XX This invention relates to novel isolated proteins and their fragments  
 XX with anti-angiogenic properties. The invention also relates to the DNA  
 XX sequences which encode the novel proteins. A wide variety of diseases are  
 XX the result of undesirable angiogenesis. The formation of new capillaries  
 XX from pre-existing vessels is essential for tumour growth and metastasis.  
 XX Basement membrane organisation is dependent on the assembly of a type IV  
 XX collagen network which may occur through the C-terminal globular non-  
 XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 XX forms are ubiquitously exhibited in human basement membranes. In the  
 XX present invention, cell surface receptors (in particular integrins) which  
 XX specifically bind anti-angiogenic proteins and peptides (in particular  
 XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 XX collagen) are disclosed. The proteins of the invention may inhibit tumour  
 XX growth, angiogenic activity in mammalian tissue or protein synthesis in  
 XX endothelial cells and thus may exhibit cytostatic activity. The DNA  
 XX sequences of the invention may be useful in gene therapy. The present  
 XX sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
 XX the invention which was derived from the amino acid sequence of the alpha  
 XX 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
 XX not appear in the specification but was created by the indexer from  
 XX information given in the specification.  
 XX Sequence 79 AA;  
 SQ  
 Query Match 89.4%; Score 126; DB 6; Length 79;  
 Best Local Similarity 92.3%; Pred. No. 8.3e-13;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27  
 DB 16 QRFTTTPFLFCNVNDVCFASRNDYS 41  
 RESULT 6  
 AAU75608  
 ID AAU75608 standard; protein; 88 AA.  
 XX  
 AC AAU75608;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.  
 XX  
 KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
 XX  
 CS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 82

/note= "Wild type Cys substituted with Ala"  
 WO200151523-A2.  
 19-JUL-2001.  
 08-JAN-2001; 2001WO-US000565.  
 07-JAN-2000; 2000US-00479118.  
 04-APR-2000; 2000US-00543371.  
 21-JUL-2000; 2000US-00625191.  
 (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 Kalluri R;  
 WPI; 2002-188037/24.  
 A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 treating disorders involving angiogenesis.  
 Claim 41; Page 153; 205pp; English.  
 The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 domain, having one or more of the characteristics selected from: (a) the  
 ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 proliferation of endothelial cells; and (c) the ability to cause  
 apoptosis of endothelial cells. Also described are the following: (1) use  
 of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 analogue or allelic variant in the preparation of a medicament for  
 treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 where the angiogenesis is mediated by one or more endothelial cell  
 integrins or one or more endothelial cell integrin subunits; or (b) by  
 promoting or inducing endothelial cell apoptosis in a tissue, where the  
 endothelial cell apoptosis is mediated by one or more endothelial cell  
 integrins or one or more endothelial cell integrin subunits; (2) use of  
 an antibody or peptide that specifically binds the alpha1, alpha2,  
 alpha3, alpha4, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the  
 preparation of a medicament for inhibiting angiogenesis or cell  
 proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 fragment or peptide of receptor-mediated angiogenesis in the preparation  
 of a medicament for treating a proliferative disease in a vertebrate,  
 where the disease is characterised by angiogenesis that is mediated by  
 receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 the presence of a medicament for promoting angiogenesis in a tissue; and  
 (5) use of integrins in the preparation of a medicament for promoting or  
 inducing angiogenesis or cell proliferation in a tissue. The fragments  
 Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 or allelic variants are useful in the preparation of a medicament for  
 treating a disorder involving inhibiting angiogenesis in a tissue, where  
 the angiogenesis is mediated by one or more endothelial cell integrins or  
 one or more endothelial cell integrin subunits; or by promoting or  
 inducing endothelial cell apoptosis in a tissue, where the endothelial  
 cell apoptosis is mediated by one or more endothelial cell integrins or  
 one or more endothelial cell integrin subunits. The medicament is useful  
 in inhibiting tumour growth and for the regression of an established  
 tumour. The present sequence represents the amino acid sequence of human  
 type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which  
 consists of residues 5-126 of Tumstatin  
 Sequence 88 AA;  
 Query Match 89.4%; Score 126; DB 5; Length 88;  
 Best Local Similarity 92.3%; Pred. No. 9.5e-13;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27  
 DB 26 QRFTTTPFLFCNVNDVCFASRNDYS 51

RESULT 7  
AAU75607  
XX AAU75607 standard; protein; 88 AA.  
XX AC  
XX AAU75607;  
XX AC  
XX 08-MAY-2002 (first entry)  
XX DT  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.  
XX XX  
XX Human; type IV collagen alpha 3 chain; cytosolic; antiangiogenic;  
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX KW Tumstatin; angiogenesis; tumour; mutant.  
XX XX  
XX OS Homo sapiens.  
XX XX  
XX FN WO200151523-A2.  
XX XX  
XX PD 19-JUL-2001.  
XX XX  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX XX  
XX PR 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX XX  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX XX  
XX PI Kalluri R;  
XX XX  
XX DR WPI; 2002-188037/24.  
XX XX  
XX XX  
XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX PT treating disorders involving angiogenesis.  
XX XX  
XX PS Claim 32; Page 152; 205pp; English.  
XX XX  
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX CC domain, having one or more of the characteristics selected from: (a) the  
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX CC proliferation of endothelial cells; and (c) the ability to cause  
XX CC apoptosis of endothelial cells. Also described are the following: (1) use  
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX CC analogue or allelic variant in the preparation of a medicament for  
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX CC where the angiogenesis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by  
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of  
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,  
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX CC preparation of a medicament for inhibiting angiogenesis or cell  
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX CC of a medicament for treating a proliferative disease in a vertebrate,  
XX CC where the disease is characterised by angiogenesis that is mediated by  
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and  
XX CC (5) use of integrins in the preparation of a medicament for promoting or  
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
XX CC or allelic variants are useful in the preparation of a medicament for  
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or  
XX CC one or more endothelial cell integrin subunits; or by promoting or  
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or  
XX CC one or more endothelial cell integrin subunits. The medicament is useful  
XX CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists  
CC of residues 45-132 of Tumstatin  
XX XX  
XX SQ Sequence 88 AA;  
XX  
XX Query Match 89,4%; Score 126; DB 5; Length 88;  
XX Best Local Similarity 92,3%; Pred. No. 9,5e-13;  
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX QY 2 QRFTTTPFLFSNVNDVGNFASRNDYS 27  
XX ||||| ||||| ||||| ||||| |||||  
XX Db 26 QRFTTTPFLFCVNDVGNFASRNDYS 51  
XX  
XX RESULT 8  
XX ADA20271  
XX ID ADA20271 standard; protein; 88 AA.  
XX AC  
XX ADA20271;  
XX AC  
XX DT 20-NOV-2003 (first entry)  
XX XX  
XX DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human;  
XX KW tumstatin 45-132.  
XX XX  
XX OS Homo sapiens.  
XX XX  
XX FN WO2003059257-A2.  
XX XX  
XX PD 24-JUL-2003.  
XX XX  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX XX  
XX PR 21-DEC-2001; 2001US-00032221.  
XX XX  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX DR  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX XX  
XX PS Claim 94; SEQ ID NO 33; 240pp; English.  
XX XX  
XX CC This invention relates to novel isolated proteins and their fragments  
XX CC with anti-angiogenic properties. The invention also relates to the DNA  
XX CC sequences which encode the novel proteins. A wide variety of diseases are  
XX CC the result of undesirable angiogenesis. The formation of new capillaries  
XX CC from pre-existing vessels is essential for tumour growth and metastasis.  
XX CC Basement membrane organisation is dependent on the assembly of a type IV  
XX CC collagen network which may occur through the C-terminal globular non-  
XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX CC forms are ubiquitously exhibited in human basement membranes. In the  
XX CC present invention, cell surface receptors (in particular integrins) which  
XX CC specifically bind anti-angiogenic proteins and peptides (in particular  
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX CC sequences of the invention may be useful in gene therapy. The present  
XX CC sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"  
XX CC protein of the invention which was derived from the amino acid sequence  
XX CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq

This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis. Basement membrane organisation is dependent on the assembly of a type IV collagen network which may occur through the C-terminal globular non-collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2 forms are ubiquitously exhibited in human basement membranes. In the present invention, cell surface receptors (in particular integrins) which specifically bind anti-angiogenic proteins and peptides (in particular the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV collagen) are disclosed. The proteins of the invention may inhibit tumour growth, angiogenic activity in mammalian tissue or protein synthesis in endothelial cells and thus may exhibit cytostatic activity. The DNA



CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of  
 CC residues 2-125 of tumstatin. Note: The present sequence is not shown in  
 CC the specification but is derived from the wild type human Tumstatin  
 CC sequence given in figure 18A (see AAU75589)

SQ Sequence 124 AA;

Query Match 89.4%; Score 126; DB 5; Length 124;  
 Best Local Similarity 92.3%; Pred. No. 1.5e-12;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27  
 |||||  
 DB 69 QRFTTTPFLFCNVNDVCFASRNDYS 94

RESULT 11

ADA20258  
 ID ADA20258 standard; protein; 124 AA.

AC ADA20258;

DT 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.

KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.

OS Homo sapiens.

PN WO2003059257-A2.

XX 24-JUL-2003.

PF 20-DEC-2002; 2002WO-US040938.

PR 21-DEC-2001; 2001US-00032221.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;

DR WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Claim 94; SEQ ID NO 20; 240pp; English.

CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"  
 CC protein of the invention which was derived from the amino acid sequence  
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (seq  
 CC ID20) does not appear in the specification but was created by the indexer  
 CC from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 89.4%; Score 126; DB 6; Length 124;  
 Best Local Similarity 92.3%; Pred. No. 1.5e-12;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTTPFLFSNVNDVSNFASRNDYS 27  
 |||||  
 DB 69 QRFTTTPFLFCNVNDVCFASRNDYS 94

RESULT 12

AAU75597  
 ID AAU75597 standard; protein; 132 AA.

AC AAU75597;

XX 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tum-2.

KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

PN WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;

XX WPI; 2002-189037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and  
 XX treating disorders involving angiogenesis.

PS Claim 31; Page 152; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI  
 CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues  
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75589)

XX Sequence 132 AA;  
 Query Match 89.4%; Score 126; DB 5; Length 132;  
 Best Local Similarity 92.3%; Pred. No. 1.6e-12;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFITMPFLFSNVNDVSNFASRNDYS 27  
 |||||  
 Db 70 QRFITMPFLFCNVNDVCFNFSRNDYS 95  
 |||||

RESULT 13  
 ADA20261  
 ID ADA20261 standard; protein; 132 AA.  
 XX  
 AC ADA20261;  
 XX

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; Ncl; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

PF 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.  
 PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCl) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCl domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.

XX Sequence 132 AA;

Query Match 89.4%; Score 126; DB 6; Length 132;  
 Best Local Similarity 92.3%; Pred. No. 1.6e-12;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFITMPFLFSNVNDVSNFASRNDYS 27  
 |||||  
 Db 69 QRFITMPFLFCNVNDVCFNFSRNDYS 94  
 |||||

RESULT 14

AAU75596  
 ID AAU75596 standard; protein; 191 AA.

XX AAU75596;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NCl domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds  
(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-40  
Perfect score: 141  
Sequence: 1 KQRTTTPFLFNSVNDVSNFASRNDYS 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141	100.0	27	14	US-10-032-221B-40
2	133	94.3	27	14	Sequence 42, Appl
3	131	92.9	27	14	Sequence 39, Appl
4	126	89.4	79	14	Sequence 26, Appl
5	126	89.4	88	14	Sequence 33, Appl
6	126	89.4	124	14	Sequence 34, Appl
7	126	89.4	124	14	Sequence 20, Appl
8	126	89.4	132	14	Sequence 23, Appl
9	126	89.4	131	14	Sequence 22, Appl
10	126	89.4	211	14	Sequence 46, Appl
11	126	89.4	211	14	Sequence 46, Appl
12	126	89.4	232	14	Sequence 304, Appl
13	126	89.4	244	14	Sequence 10, Appl
14	110	78.0	229	14	Sequence 306, Appl
15	110	78.0	309	9	US-09-925-297-496

16	109	77.3	229	14	US-10-206-699-302
17	109	77.3	229	14	US-10-032-221B-2
18	109	77.3	406	9	US-09-925-302-507
19	109	77.3	1669	15	US-10-372-683-8
20	105	74.5	22	14	US-10-206-699-266
21	105	74.5	25	14	US-10-032-221B-37
22	97	68.8	25	14	US-10-032-221B-38
23	95	67.4	22	14	US-10-206-699-265
24	95	67.4	1759	15	US-10-369-493-7032
25	93	66.0	22	14	US-10-206-699-267
26	90	63.8	20	14	US-10-206-699-289
27	90	63.8	20	14	US-10-032-221B-29
28	87	61.7	46	9	US-09-864-761-48095
29	87	61.7	1744	15	US-10-369-493-5832
30	85	60.3	142	9	US-09-864-761-38021
31	85	60.3	228	14	US-10-206-699-307
32	84	59.6	18	14	US-10-206-699-254
33	84	59.6	18	14	US-10-206-699-260
34	82	58.2	227	14	US-10-206-699-303
35	82	58.2	227	14	US-10-032-221B-6
36	82	58.2	430	9	US-09-925-302-518
37	82	58.2	459	15	US-10-331-496A-27
38	82	58.2	459	15	US-10-372-683-30
39	82	58.2	1712	10	US-09-961-403-9
40	78	55.3	18	14	US-10-206-699-259
41	78	55.3	22	14	US-10-206-699-270
42	76	53.9	18	14	US-10-206-699-261
43	75	53.2	22	14	US-10-206-699-268
44	74	52.5	18	14	US-10-206-699-253
45	74	52.5	20	14	US-10-206-699-290

## ALIGNMENTS

### RESULT 1

US-10-032-221B-40  
; Sequence 40, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghubram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 140-1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 40  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8-3)  
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine residue at position 79 and 85)  
; OTHER INFORMATION: residues at positions 79 and 85)  
US-10-032-221B-40

Query Match 100.0%; Score 141; DB 14; Length 27;

```
Best Local Similarity 100.0%; Pred. No. 4.2e-15; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0;

Qy 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||
Db 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||

RESULT 2
US-10-032-221B-42
; Sequence 42, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ed for the leucine residue at position 68 of the full-length Tumst
; OTHER INFORMATION: tatin molecule, and aspartic acid has been substituted for the cy
; OTHER INFORMATION: steine residues at positions 79 and 85)
US-10-032-221B-42

Query Match 94.3%; Score 133; DB 14; Length 27;
Best Local Similarity 92.6%; Pred. No. 7.6e-14;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||
Db 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||

RESULT 3
US-10-032-221B-39
; Sequence 39, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26

Query Match 89.4%; Score 126; DB 14; Length 79;
Best Local Similarity 92.3%; Pred. No. 3.3e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTNPFLFSNVNDSNFASRNDYS 27
    |||||
Db 16 QRTTNPFLFSNVNDSNFASRNDYS 41
    |||||

RESULT 5
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
```

```
PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T8 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted
; OTHER INFORMATION: d for the leucine residue at position 68 of the full-length Tumst
; OTHER INFORMATION: atin molecule)
US-10-032-221B-39

Query Match 92.9%; Score 131; DB 14; Length 27;
Best Local Similarity 92.6%; Pred. No. 1.6e-13;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||
Db 1 KQRTTNPFLFSNVNDSNFASRNDYS 27
    |||||

RESULT 4
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-5 (amino acids 54-132 of SEQ ID NO:10)
US-10-032-221B-26

Query Match 89.4%; Score 126; DB 14; Length 79;
Best Local Similarity 92.3%; Pred. No. 3.3e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTNPFLFSNVNDSNFASRNDYS 27
    |||||
Db 16 QRTTNPFLFSNVNDSNFASRNDYS 41
    |||||

RESULT 5
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match      89.4%; Score 126; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 3.7e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFSNVNDSNFASRNDYS 27
DB      25 QRFTHMPFLFCNVNDCVNFASRNDYS 50

RESULT 6
US-10-032-221B-34
; Sequence 34, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine
OTHER INFORMATION: has been substituted for the cysteine residue at position 125 of
OTHER INFORMATION: the full-length Tumstatin molecule)

```

```

US-10-032-221B-34

Query Match      89.4%; Score 126; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 3.7e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFSNVNDSNFASRNDYS 27
DB      25 QRFTHMPFLFCNVNDCVNFASRNDYS 50

RESULT 7
US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match      89.4%; Score 126; DB 14; Length 124;
Best Local Similarity 92.3%; Pred. No. 5.5e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFSNVNDSNFASRNDYS 27
DB      69 QRFTHMPFLFCNVNDCVNFASRNDYS 94

RESULT 8
US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224

```

```
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 23
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)
US-10-032-221B-23

Query Match      89.4%; Score 126; DB 14; Length 132;
Best Local Similarity 92.3%; Pred. No. 6e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRFTHMPFLFSNVNDVSNFASRNDYS 27
      ||||| ||||| ||||| ||||| |||||
Db      69 QRFTHMPFLFCNVNDVSNFASRNDYS 94

RESULT 9
US-10-032-221B-22
; Sequence 22, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)
US-10-032-221B-22

Query Match      89.4%; Score 126; DB 14; Length 191;
Best Local Similarity 92.3%; Pred. No. 9.1e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRFTHMPFLFSNVNDVSNFASRNDYS 27
      ||||| ||||| ||||| ||||| |||||
Db      16 QRFTHMPFLFCNVNDVSNFASRNDYS 41

RESULT 10
US-10-032-221B-21
; Sequence 46, Application US/10270877
; Publication No. US20030049791A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
```

```
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match      89.4%; Score 126; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRFTHMPFLFSNVNDVSNFASRNDYS 27
      ||||| ||||| ||||| ||||| |||||
Db      69 QRFTHMPFLFCNVNDVSNFASRNDYS 94

RESULT 11
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match      89.4%; Score 126; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 QRFTHMPFLFSNVNDVSNFASRNDYS 27
      ||||| ||||| ||||| ||||| |||||
Db      69 QRFTHMPFLFCNVNDVSNFASRNDYS 94

RESULT 12
US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
```



; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 304  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 3 chain  
US-10-206-699-304

Query Match 89.4%; Score 126; DB 14; Length 244;  
Best Local Similarity 92.3%; Pred. No. 1.1e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27  
|||||  
DB 57 QRFTHMPFLFCNVNVCNFASRNDYS 82  
|||||

RESULT 13  
US-10-032-221B-10  
; Sequence 10, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram

; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-10

Query Match 89.4%; Score 126; DB 14; Length 244;  
Best Local Similarity 92.3%; Pred. No. 1.1e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27  
|||||  
DB 69 QRFTHMPFLFCNVNVCNFASRNDYS 94  
|||||

RESULT 14  
US-10-206-699-306  
; Sequence 306, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.

; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MEHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 306  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 5 chain  
US-10-206-699-306

Query Match 78.0%; Score 110; DB 14; Length 229;  
Best Local Similarity 73.1%; Pred. No. 3.7e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27  
|||||  
DB 55 RRFSTMPFMCNINNVCFASRNDYS 80  
|||||

RESULT 15  
US-09-925-297-496  
; Sequence 496, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 496  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (247)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-496

Query Match 78.0%; Score 110; DB 9; Length 309;  
Best Local Similarity 73.1%; Pred. No. 5.2e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFSNVDSNFASRNDYS 27  
|||||  
DB 135 RRFSTMPFMCNINNVCFASRNDYS 160  
|||||

Search completed: April 5, 2004, 07:36:07  
Job time : 17.9322 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221b-40

Perfect score: 141

Sequence: 1 QRFTHMPFLFNVNDVNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*

1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*

2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*

3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*

4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*

5: /cgn2\_6/prodata/2/iaa/PCUTUS COMB.pep.\*

6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	89.4	211	4	US-09-512-563C-46
2	126	89.4	218	2	US-08-399-889-25
3	126	89.4	218	3	US-09-167-364-25
4	126	89.4	218	3	US-09-439-897-4
5	126	89.4	268	4	US-09-589-927-6
6	126	89.4	268	4	US-09-277-665-6
7	126	89.4	268	4	US-09-589-987-6
8	125	88.7	471	2	US-08-399-889-24
9	125	88.7	471	3	US-09-167-364-24
10	125	88.7	471	3	US-09-439-897-2
11	110	78.0	264	4	US-09-589-927-10
12	110	78.0	264	4	US-09-277-665-10
13	110	78.0	264	4	US-09-589-987-10
14	109	77.3	260	4	US-09-589-927-2
15	109	77.3	260	4	US-09-277-665-2
16	109	77.3	260	4	US-09-589-987-2
17	85	60.3	260	4	US-09-589-927-12
18	85	60.3	260	4	US-09-277-665-12
19	85	60.3	260	4	US-09-589-987-12
20	82	58.2	258	4	US-09-589-927-4
21	82	58.2	258	4	US-09-277-665-4
22	82	58.2	258	4	US-09-589-987-4
23	68	48.2	260	4	US-09-589-927-8
24	68	48.2	260	4	US-09-277-665-8
25	68	48.2	260	4	US-09-589-987-8
26	48	34.0	1694	1	US-08-494-168-2
27	46	32.6	107	3	US-09-102-528-23

28	46	32.6	107	3	US-09-102-528-27	Sequence 27, Appl
29	46	32.6	286	4	US-08-252-991A-20515	Sequence 20515, A
30	46	32.6	587	3	US-09-102-528-30	Sequence 30, Appl
31	46	32.6	736	3	US-09-102-528-29	Sequence 29, Appl
32	45	31.9	1117	4	US-09-252-991A-23416	Sequence 23416, A
33	44	31.2	167	5	PCT-US95-13813-9	Sequence 9, Appl
34	44	31.2	476	4	US-09-339-159B-4	Sequence 4, Appl
35	44	31.2	493	3	US-09-198-956-10	Sequence 10, Appl
36	44	31.2	493	3	US-09-198-955A-12	Sequence 12, Appl
37	44	31.2	493	4	US-09-594-531-12	Sequence 12, Appl
38	44	31.2	493	4	US-09-670-141-10	Sequence 10, Appl
39	44	31.2	493	4	US-10-072-152-12	Sequence 12, Appl
40	44	31.2	793	2	US-08-468-558-5	Sequence 5, Appl
41	44	31.2	793	3	US-08-676-444-5	Sequence 11, Appl
42	43.5	30.9	288	2	US-08-424-641B-11	Sequence 11, Appl
43	43.5	30.9	288	2	US-08-820-980-11	Sequence 11, Appl
44	43.5	30.9	288	2	US-08-826-439-11	Sequence 11, Appl
45	43	30.5	192	4	US-09-489-039A-13105	Sequence 13105, A

## ALIGNMENTS

RESULT 1  
US-09-512-563C-46  
; Sequence 46, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512,563C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-09-512-563C-46

Query Match 89.4%; Score 126; DB 4; Length 211;  
Best Local Similarity 92.3%; Pred. No. 3.3e-12;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFNVNDVNFASRNDYS 27  
DB 69 QRFTHMPFLFNVNDVNFASRNDYS 94

RESULT 2  
US-08-399-889-25  
; Sequence 25, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 218  
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match      89.4%; Score 126; DB 2; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
    |||||
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68
    |||||

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      89.4%; Score 126; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
    |||||
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68
    |||||

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      89.4%; Score 126; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. No. 3.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
    |||||
Db 43 QRTTMPFLFCNVNDVCFASRNDYS 68
    |||||

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match      89.4%; Score 126; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
    |||||
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118
    |||||

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match      89.4%; Score 126; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. No. 4.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFSNVNDVSNFASRNDYS 27
    |||||
Db 93 QRTTMPFLFCNVNDVCFASRNDYS 118
    |||||

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match      89.4%; Score 126; DB 4; Length 268;
```



Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27  
:|||||:|:|:|||||  
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 13  
US-09-589-987-10  
; Sequence 10, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-10

Query Match 78.0%; Score 110; DB 4; Length 264;  
Best Local Similarity 73.1%; Pred. No. 1.4e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27  
:|||||:|:|:|||||  
Db 90 RRFSTMPFMCNINNVCFASRNDYS 115

RESULT 14  
US-09-589-927-2  
; Sequence 2, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-2

Query Match 77.3%; Score 109; DB 4; Length 260;  
Best Local Similarity 73.1%; Pred. No. 2e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27  
:|||||:|:|:|||||  
Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

RESULT 15  
US-09-277-665-2  
; Sequence 2, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-2

Query Match 77.3%; Score 109; DB 4; Length 260;  
Best Local Similarity 73.1%; Pred. No. 2e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFSNVNDSNFASRNDYS 27  
:|||||:|:|:|||||  
Db 86 RRFSTMPFMCNINNVCFASRNDYS 111

Search completed: April 5, 2004, 07:07:25  
Job time : 6.3414 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 3.91041 Seconds  
(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110  
Sequence: 1 LFCNVNVCNCFASRNDYS 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 78: \*  
1: Pirl: \*  
2: Pirl: \*  
3: Pirl: \*  
4: Pirl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	84.5	220	B49736	collagen alpha 3(I)
2	93	84.5	1670	CGHU3B	collagen alpha 3(I)
3	92	83.6	161	S49488	collagen alpha 3(I)
4	92	83.6	246	I48302	collagen alpha 3(I)
5	92	83.6	258	B61228	collagen alpha 3(I)
6	92	83.6	471	A39024	collagen alpha 3(I)
7	92	83.6	1669	CGHU4B	collagen alpha 1(I)
8	92	83.6	1669	CGM34B	collagen alpha 1(I)
9	90	81.8	253	I48304	collagen alpha 5(I)
10	90	81.8	754	A55267	collagen alpha 5(I)
11	90	81.8	1691	S22917	collagen alpha 3(I)
12	80	72.7	1752	A45407	collagen alpha 3(I)
13	76	69.1	1747	A54121	collagen alpha-4 c
14	75	69.1	1763	S16366	collagen alpha 2(I)
15	75	68.2	1744	S40391	collagen alpha 1(I)
16	71	64.5	1758	T29350	hypothetical prote
17	71	64.5	1759	T29351	collagen alpha 2(I)
18	67	60.9	261	A34476	collagen alpha 2(I)
19	64	58.2	1691	CGHU6B	collagen alpha 6(I)
20	59	53.6	312	I48303	collagen alpha 4(I)
21	59	53.6	623	A45137	collagen alpha 4(I)
22	59	53.6	1690	CGHU2B	collagen alpha 4(I)
23	59	53.6	1775	A31893	collagen alpha 1(I)
24	58	52.7	453	S18804	collagen alpha 4(I)
25	58	52.7	775	A61228	collagen alpha 2(I)
26	58	52.7	1707	A33526	collagen alpha 2(I)
27	58	52.7	1712	CGHU2B	collagen alpha 2(I)
28	53	48.2	79	C43928	probable collagen
29	43	44.5	346	T46914	hypothetical prote

30 49 44.5 438 2 E84579  
31 49 44.5 1761 2 T13990  
32 47 42.7 1477 2 T18534  
33 46.5 42.3 419 2 S41607  
34 46 41.8 257 2 H75419  
35 46 41.8 338 2 T05036  
36 46 41.8 760 2 T41644  
37 45 40.9 840 2 T38528  
38 43 39.1 114 2 T18089  
39 43 39.1 1270 2 T51227  
40 43 39.1 2946 2 T00867  
41 42.5 38.6 553 2 B72863  
42 42.5 38.6 578 2 S50446  
43 42.5 38.6 869 2 P97126  
44 42.5 38.6 1270 2 T26720  
45 42 38.2 75 2 JC6048

## ALIGNMENTS

### RESULT 1

B49736  
collagen alpha 3(IV) chain, medium splice form - human (fragment)

N;Contains: collagen alpha 3(IV) chain, splice form GP-V

C;Species: Homo sapiens (man)

C;Date: 03-May-1994 #sequence revision 12-Nov-1999 #text\_change 17-Mar-2000

C;Accession: B49736; D49736; S69111

R;Feng, L.; Xia, Y.; Wilson, C.B.

J. Biol. Chem. 269, 2342-2348, 1994

A;Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.

A;Reference number: A49736; MUID:94124597; PMID:8294492

A;Accession: B49736

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 169-220 <FEN1>

A;Accession: D49736

A;Status: nucleic acid sequence not shown; translation not shown

A;Residues: 22-220 <FEN2>

A;Cross-references: GB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107

A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank

R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Presquet, V.J.; Cervera, J.; W

Eur. J. Biochem. 229, 754-760, 1995

A;Title: Characterization and expression of multiple alternatively spliced transcripts

into antigen and one of its alternative forms.

A;Reference number: S69111; MUID:95278230; PMID:7758473

A;Accession: S69111

A;Molecule type: mRNA

A;Residues: 1-45,169-204,'L',206-220 <PEN>

C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.

C;Genetics:

A;Gene: GDB:COL4A3

A;Cross-references: GDB:128351; OMIM:120070

A;Map position: 2q36-2q37

C;Superfamily: collagen alpha 1(IV) chain

C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrac

F;1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pre

F;1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #statu

F;22-220/Domain: carboxyl-terminal nonhelical, NCI <NC1>

F;22-220/Domain: collagen IV carboxyl-terminal repeat <Ctrl>

F;34-134/Domain: collagen IV carboxyl-terminal repeat <Ctrl>

Query Match 84.5%; Score 93; DB 2; Length 220;

Best Local Similarity 94.4%; Pred. No. 2,7e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

DB 86 LFCNVNVCNCFASRNDYS 103

### RESULT 2

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human  
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
 C/Species: Homo sapiens (man)  
 C/Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
 C/Accession: A54763; A43928; A44043; A45971; A39786  
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.  
 J. Biol. Chem. 269, 23013-23017, 1994  
 A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
 A/Reference number: A54763; MUID:94364994; PMID:8083201  
 A/Accession: A54763  
 A/Molecule type: mRNA  
 A/Residues: 1-1670 <VAR>  
 A/Cross-references: GB:X80031; NID:g577563; PID:g577564  
 A/Experimental source: Kidney  
 R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
 J. Clin. Invest. 89, 592-601, 1992  
 A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha  
 A/Reference number: A43928; MUID:92147878; PMID:1737849  
 A/Accession: A43928  
 A/Molecule type: mRNA  
 A/Residues: 1331-1524, 'I', 1526-1670 <TUR>  
 A/Cross-references: GB:M81379  
 A/Experimental source: Kidney  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 267, 19780-19784, 1992  
 A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture  
 cation.  
 A/Reference number: A44043; MUID:93015826; PMID:1400291  
 A/Accession: A44043  
 A/Molecule type: DNA; mRNA  
 A/Residues: 1386-1670 <QUI>  
 A/Cross-references: GB:M92993; NID:gi77895; PID:AAA21610.1; PID:gi77896  
 A/Note: sequence extracted from NCBI backbone (NCBIP:115597)  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 269, 17358, 1994  
 A/Reference number: A44738; MUID:94274734; PMID:8006044  
 A/Contents: annotation; erratum; correction to intronic sequence in A44043  
 R/Bernal, D.; Quinones, S.; Saus, J.  
 J. Biol. Chem. 268, 12090-12094, 1993  
 A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
 A/Reference number: A45971; MUID:93280184; PMID:8505332  
 A/Accession: A45971  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1427-1444 <BER>  
 A/Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly ident  
 R/Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Reiders, S.T.  
 Am. J. Hum. Genet. 49, 545-554, 1991  
 A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
 A/Reference number: A39786; MUID:91353570; PMID:1882840  
 A/Accession: A39786  
 A/Molecule type: mRNA  
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>  
 A/Cross-references: GB:S55790; NID:g234418; PID:BA819637.1; PID:g234419  
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
 C/Genetics:  
 A/Gene: GDB:COL4A3  
 A/Cross-references: GDB:128351; OMIM:120070  
 A/Map position: 2q36-2q37  
 A/Introns: 1395/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete  
 A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
 er associations in the interrupted helical domain (with disulfide and desmosine cross-li  
 C/Function:  
 A/Description: minor structural component of extracellular basement membrane in kidney  
 A/Superfamily: collagen alpha 1(IV) chain  
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>  
 F:29-42/Domain: amino-terminal nonhelical, NH1 <NHL>

F:43-1438/Region: interrupted helical  
 F:791-793/Region: cell attachment (R-G-D) motif  
 F:996-998/Region: cell attachment (R-G-D) motif  
 F:1154-1156/Region: cell attachment (R-G-D) motif  
 F:1306-1308/Region: cell attachment (R-G-D) motif  
 F:1345-1347/Region: cell attachment (R-G-D) motif  
 F:1432-1434/Region: cell attachment (R-G-D) motif  
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NCL <NCL>  
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>  
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>  
 F:31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred:  
 F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F:1505-1511.1616-1622/Disulfide bonds: #status predicted  
 F:1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
 Query Match 84.5%; Score 93; DB 1; Length 1670;  
 Best Local Similarity 94.4%; Pred. No. 1.4e-05;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LFCNVNVCNCFASNDYS 19  
 DB 1503 LFCNVNVCNCFASNDYS 1520  
 RESULT 3  
 S49488  
 collagen alpha 3(IV) chain - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999  
 C/Accession: S49488  
 R/Oberbaumer, I.  
 A/Description: Cloning of the NCL domains fo the minor collagen IV chains of mouse via  
 cells.  
 A/Reference number: S49487  
 A/Accession: S49488  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-161 <OBE>  
 A/Cross-references: EMBL:X82205; NID:g559472; PID:CAA57689.1; PID:g555916  
 C/Superfamily: collagen alpha 1(IV) chain  
 Query Match 83.6%; Score 92; DB 2; Length 161;  
 Best Local Similarity 88.9%; Pred. No. 2.9e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LFCNVNVCNCFASNDYS 19  
 DB 12 LFCNVNVCNCFASNDYS 29  
 RESULT 4  
 I48302  
 collagen alpha 3(IV) chain - mouse (fragment)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
 C/Accession: I48302; S47278  
 R/Miner, J.H.; Sanes, J.R.  
 J. Cell Biol. 127, 879-891, 1994  
 A/Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
 A/Reference number: A54979; MUID:95050957; PMID:7962085  
 A/Accession: I48302  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-246 <RES>  
 A/Cross-references: EMBL:Z35166; NID:g535197; PID:g535198  
 C/Superfamily: collagen alpha 1(IV) chain  
 Query Match 83.6%; Score 92; DB 2; Length 246;  
 Best Local Similarity 88.9%; Pred. No. 4.1e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



C; Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; F; 1-239/Domain: collagenous (fragment) #status predicted <COL> F; 239-471/Domain: carboxy-terminal nonhelical, NC1 #status predicted <NC1> F; 239-353/Domain: repeat NC1 #status predicted <NC1> F; 354-471/Domain: repeat NC1 #status predicted <NC12> F; 232-238/Modified site: hydroxyproline (Pro) #status experimental F; 306-312, 417-423/disulfide bonds: #status predicted

Query Match 83.6%; Score 92; DB 2; Length 471;  
Best Local Similarity 88.9%; Pred. No. 6.9e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNFAASRNDYS 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 304 LFCNINDVCNFAASRNDYS 321

RESULT 7  
CGHU4B  
collagen alpha 1(IV) chain precursor - human  
N; Alternate names: procollagen alpha 1(IV) chain  
C; Species: Homo sapiens (man)  
C; Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C; Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A.  
R; Soinenen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A; Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen  
A; Reference number: S16876; MUID:89340433; PMID:2701944  
A; Accession: S16876  
A; Status: nucleic acid sequence not shown; translation not shown  
A; Molecule type: DNA  
A; Residues: 1-1669 <SOIL>  
A; Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803  
A; Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R; Soinenen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A; Title: The structural genes for alpha 1 and alpha 2 chains of human type IV collagen  
A; Reference number: A92690; MUID:89034231; PMID:3182844  
A; Accession: A32117  
A; Molecule type: DNA  
A; Residues: 1-28 <SOI2>  
A; Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R; Roesschl, E.; Follmer, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A; Title: The genes for the alpha(IV) and alpha2(IV) chains of human basement membrane  
A; Reference number: S02738; MUID:89030632; PMID:2846280  
A; Accession: S02738  
A; Status: translation not shown  
A; Molecule type: DNA  
A; Residues: 1-6, 'L', 8-28 <POB>  
A; Cross-references: EMBL:X12784; NID:G30072  
R; Brazel, D.; Oberbaumer, I.; Bieringer, H.; Babel, W.; Glangville, R.W.; Deutzmann, R.  
Eur. J. Biochem. 168, 529-536, 1987  
A; Title: Completion of the amino acid sequence of the alpha 1 chain of human basement membrane  
A; Reference number: S00048; MUID:88029471; PMID:3311751  
A; Accession: S00048  
A; Molecule type: mRNA  
A; Residues: 1-318, 'A', 320-944 <BRAL>  
A; Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A; Accession: S25826  
A; Molecule type: protein  
A; Residues: 271-319, 'A', 320-554 <BRA2>  
R; Glangville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A; Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7  
A; Reference number: A23115; MUID:86004708; PMID:4043082  
A; Accession: A23115  
A; Molecule type: protein  
A; Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A; Experimental source: Placenta  
R; Soinenen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1987

[illegible]

A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
 A;Reference number: A28066; MUID:88243724; PMID:3379041  
 A;Accession: A28066  
 A;Molecule type: mRNA  
 A;Residues: 1-129 <X1>  
 A;Cross-references: EMBL:J03758; NID:G192669; PIDN:AAA37439.1; PID:G192670  
 R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985  
 A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
 A;Reference number: A02864; MUID:85127033; PMID:2578961  
 A;Accession: A02864  
 A;Molecule type: mRNA  
 A;Residues: 1276-1669 <OB>  
 A;Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876  
 R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G. Gene 43, 301-304, 1986  
 A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox  
 A;Reference number: A25636; MUID:86301886; PMID:3755692  
 A;Accession: A25636  
 A;Molecule type: mRNA  
 A;Residues: 1149-1396, S', 1398-1424 <NAT>  
 A;Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287  
 A;Note: the authors translated the codon CAG for residue 1374 as Arg  
 R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla  
 J. Biol. Chem. 262, 8496-8499, 1987  
 A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
 A;Reference number: A94680; MUID:87250460; PMID:3597383  
 A;Accession: A29301  
 A;Molecule type: mRNA  
 A;Residues: 1441-1669 <XUR>  
 A;Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G987115  
 R;Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y. J. Biol. Chem. 263, 12310-12314, 1988  
 A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc  
 A;Reference number: S19079; MUID:88315019; PMID:2842328  
 A;Accession: S19079  
 A;Molecule type: DNA  
 A;Residues: 1-28 <K12>  
 A;Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G466503  
 R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G. J. Biol. Chem. 263, 19274-19277, 1988  
 A;Title: Head-to-head arrangement of murine type IV collagen genes.  
 A;Reference number: A92702; MUID:89066738; PMID:3198626  
 A;Accession: A32003  
 A;Molecule type: DNA  
 A;Residues: 1-28 <KAY>  
 A;Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449  
 R;Burbello, P.D.; Martin, G.R.; Yamada, Y. Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
 A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
 A;Reference number: A94220; MUID:89071759; PMID:3200951  
 A;Accession: A31766  
 A;Molecule type: DNA  
 A;Residues: 1-28 <BUR>  
 A;Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668  
 R;Sakurai, Y.; Sullivan, M.; Yamada, Y. J. Biol. Chem. 261, 6654-6657, 1986  
 A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
 A;Reference number: S19094; MUID:86196099; PMID:3009468  
 A;Accession: S19094  
 A;Molecule type: DNA  
 A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>  
 A;Cross-references: EMBL:M13027  
 R;Schuppan, D.; Timpi, R.; Glanville, R.W. FEBS Lett. 115, 297-300, 1980  
 A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty  
 A;Reference number: S16909; MUID:80246483; PMID:6772473  
 A;Accession: S16909  
 A;Molecule type: protein  
 A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123  
 R;Schuppan, D.; Glanville, R.W.; Timpi, R. Eur. J. Biochem. 123, 505-512, 1982  
 A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amid

A;Reference number: A25991; MUID:82186723; PMID:6804236  
 A;Accession: A25991  
 A;Molecule type: protein  
 A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 1  
 61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-111  
 A;Accession: B25991  
 A;Molecule type: protein  
 A;Residues: 1173-1187, 'X', 1189-1187, 'X', 1189-1205, 'O', 1207, 'XE', 1210-123  
 3, 'SP', 1266, 'IV', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 13  
 R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpi, R. Eur. J. Biochem. 139, 401-410, 1984  
 A;Title: Subunit structure and assembly of the globular domain of basement-membrane co  
 A;Reference number: S17801; MUID:84132058; PMID:6698021  
 A;Accession: S17801  
 A;Molecule type: protein  
 A;Residues: 1435-1443 <WEB>  
 C;Genetics:  
 A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3  
 A;Note: the list of introns may be incomplete  
 C;Superfamily: collagen alpha 1(IV) chain  
 C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular m  
 F;1-27/Dmain: signal sequence #status predicted <SIG>  
 F;128-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
 F;163-1440/Dmain: 7S <7SD>  
 F;163-1440/Dmain: collagenous, triple helix <COL>  
 F;597-599/Region: cell attachment (R-G-D) motif  
 F;781-783/Region: cell attachment (R-G-D) motif  
 F;917-919/Region: cell attachment (R-G-D) motif  
 F;968-970/Region: cell attachment (R-G-D) motif  
 F;1441-1669/Dmain: carboxyl-terminal nonhelical, NCL <NCL>  
 F;1441-1552/Region: duplication  
 F;1553-1669/Region: duplication  
 F;31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted  
 F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F;971.974.977.986.989.1001.1007.1019.1022.1031.1037.1040.1055.1060.1063.1075.1078.1090  
 92.1298.1310.1313.1322.1337.1346.1349.1422.1425.1431.1437.1440/Modified site: hydroxyp  
 F;1214.1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F;1505-1511.1616-1622/Disulfide bonds: #status predicted

Query Match 83.68; Score 92; DB 1; Length 1669;  
 Best Local Similarity 88.98; Pred. No. 1.9e-05;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASNDYS 19  
 |||||:|||||  
 DB 1503 LFCNVNVCNCFASNDYS 1520

RESULT 9  
 I48304  
 collagen alpha 5(IV) chain - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999  
 C;Accession: I48304; S47280  
 R;Miner, J.H.; Sames, J.R. J. Cell Biol. 127, 879-891, 1994  
 A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: seq  
 A;Reference number: A54979; MUID:95050957; PMID:7962065  
 A;Accession: I48304  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-253 <RES>  
 A;Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202  
 C;Superfamily: collagen alpha 1(IV) chain

Query Match 81.88; Score 90; DB 2; Length 253;  
 Best Local Similarity 83.38; Pred. No. 8.1e-06;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASNDYS 19  
 |||||:|||||  
 DB 87 MFCNVNVCNCFASNDYS 104

## RESULT 10

A55267  
collagen alpha 5(IV) chain - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999  
C:Accession: A55267  
R:Zheng, K.; Thorne, P.S.; Marram, P.; Bauman, R.; McInnes, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-1993, 1994  
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked type IV.  
A:Reference number: A55267; MUID:94224868; PMID:8171024  
A:Accession: A55267  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-754 <ZHE>  
A:Cross-references: GB:U07888; NID:G469547; PIDN:AA60258.1; PID:G469548  
C:Superfamily: collagen alpha 1(IV) chain

Query Match	81.8%	Score 90;	DB 2;	Length 754;
Best Local Similarity	83.3%	Pred. No. 2e-05;		
Matches	15;	Conservative	2;	Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCVFASNDYS 19  
DB 595 MFCNNVNVCFASNDYS 612.

## RESULT 11

S22917  
collagen alpha 5(IV) chain precursor, renal splice form - human  
N:Alternate names: procollagen alpha 5(IV) chain  
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 27-Feb-1997 #text\_change 21-Jul-2000  
C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35  
R:Zhong, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identification of Alport syndrome patient.  
A:Reference number: S22917; MUID:92316923; PMID:1352287  
A:Accession: S22917  
A:Molecule type: mRNA  
A:Residues: 1-967 <ZHO>  
A:Cross-references: GB:M90464; NID:G180826; PIDN:AA52046.1; PID:G553234  
R:Zhong, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A:Title: Structure of the human type IV collagen COL4A5 gene.  
A:Reference number: A54365; MUID:94165049; PMID:8120014  
A:Accession: A54365  
A:Molecule type: DNA  
A:Residues: 1-922 <ZHZ>  
A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AA27816.1; PID  
R:Zhong, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggvason, K.  
Science 261, 1167-1169, 1993  
A:Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm  
A:Reference number: A57079; MUID:93361972; PMID:8356449  
A:Accession: A57079  
A:Molecule type: DNA  
A:Residues: 1-27 <ZHA>  
A:Cross-references: GB:237153; NID:G587203; PIDN:CAA85512.1; PID:G587204  
R:PhilaJanlemi, T.; Pohjola, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin  
A:Reference number: A37122; MUID:90337990; PMID:2380186  
A:Accession: A37122  
A:Molecule type: mRNA  
A:Residues: 84-439, 'LALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>  
A:Cross-references: GB:J05558; EMBL:M58526; NID:G1314209  
A:Note: submitted to the EMBL Data Library, February 1991  
A:Note: the authors translated the codon GCC for residue 115 as Val  
R:Renieri, A.; Seri, M.; Myers, J.C.; PhilaJanlemi, T.; Massella, L.; Rizzoni, G.; De Ma

```

F11534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F11644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F1129/Motified site: allylsine (Lys) #status predicted

Query Match          72.7%;   Score 80; DB 2; Length 1752;
Best Local Similarity 77.8%;   Pred. No. 0.0011;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 LFCNVNVCVNFASGRNDYS 19
|||||:|:|:||||||
Db       1586 LFCINNVCVHVRNDYS 1603

RESULT 13
A$4121
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
NIAlternate names: collagen alpha 2(IV) chain homolog
NCISpecies: Strongylocentrotus purpuratus (purple urchin)
C$Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C$Accession: A54121; S44317
R$Expósito: J.Y.; Suzuki, H.; Geourion, C.; Garrone, R.; Solursh, M.; Ramirez,
J.R. Biol. Chem. 269, 13167-13171, 1994
A$Title: Identification of a cell lineage-specific gene coding for a sea urchin
A$Reference number: A54121; MID:94230414; PMID:8175744
A$Accession: A54121
A$Molecule type: mRNA
A$Residues: 1-1747 <EXP>
A$Cross-references: EMBL:X76730; NID:g483606; PIDN:CAA54146.1; PID:g483607
C$Genetics:
A$Gene: COLP4alpha
C$Superfamily: collagen alpha 1(IV) chain

Query Match          69.1%;   Score 76; DB 2; Length 1747;
Best Local Similarity 77.8%;   Pred. No. 0.0041;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      2 LFCNVNVCVNFASGRNDYS 19
|||||:|:|:|||||
Db       1584 LFCFNVCVNYASGRNDRS 1601

RESULT 14
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C$Species: Ascaris suum (pig roundworm)
C$Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C$Accession: S16366
R$Petitt, J.; Kingston, I.B.
J. Biol. Chem. 268, 16149-16156, 1991
A$Title: The complete primary structure of a nematode alpha-2(IV) collagen and
A$Reference number: S16366; MID:91340768; PMID:1714907
A$Accession: S16366
A$Molecule type: mRNA
A$Residues: 1-1763 <JB1>
A$Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C$Genetics:
A$Introns: 229/3; 266/3; 305/3; 350/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1;
C$Superfamily: collagen alpha 1(IV) chain
C$Keywords: alternative splicing; basement membrane; cell binding; coiled coil;
F1-26/Domain: signal sequence #status predicted <SIG>
F1-2763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F1-27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F1-43-1529/Domain: collagenous #status predicted <COL>
F1-97-199/Region: cell attachment (R-G-D) motif
F1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F1530-1638/Domain: repeat NC1 #status predicted <NC11>
F1639-1763/Domain: repeat NC1 #status predicted <NC12>
F1-34-39, 41, 536-539/Disulfide bonds: interchain #status predicted
F126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F1593-1599, 1702-1709/Disulfide bonds: #status predicted

Query Match          69.1%;   Score 76; DB 2; Length 1763;
Best Local Similarity 77.8%;   Pred. No. 0.0041;

```

Mon Apr 5 07:53:16 2004

Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFSRNDYS 19  
 |||:|||||:|||||  
 Db 1591 LFCNVNVCNFSRNDYS 1608

## RESULT 15

S40991  
 collagen alpha 1(IV) chain precursor - Caenorhabditis elegans  
 N:Alternate names: protein K04H4.1  
 C:Species: Caenorhabditis elegans  
 C:Date: 03-May-1994 #sequence revision 02-Aug-1994 #text\_change 13-Aug-1999  
 C:Accession: S40991; S44442; S13651; B34476  
 R:Ainscough, R.  
 submitted to the EMBL Data Library, October 1993  
 A:Reference number: S40991  
 A:Accession: S40991  
 A:Molecule type: DNA  
 A:Residues: 1-1744 <AIN>  
 A:Cross-references: EMBL:Z27078; NID:g414627; PID:g414628  
 R:Kramer, J.M.  
 submitted to the EMBL Data Library, December 1990  
 A:Reference number: S44442  
 A:Accession: S44442  
 A:Molecule type: DNA  
 A:Residues: 1-129, 'GFGMPGLAGPQGSGQNGNPGRLSGPPGGVNSQGRKGVKSGSRGVPLP', 209-281, 'PV  
 15', 'D', '817-1260', 'P', '1262-1515', 'P', '1709-1744' <KRA>  
 A:Cross-references: EMBL:X56979; NID:g6675; PIDN:CAA40299.1; PID:g6676  
 R:Guo, X.; Johnson, J.J.; Kramer, J.M.  
 Nature 349, 707-709, 1991  
 A:Title: Embryonic lethality caused by mutations in basement membrane collagen of C. ele  
 A:Reference number: S13651; MUID:91141582; PMID:1996137  
 A:Accession: S13651  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-129, 'GFGMPGLAGPQGSGQNGNPGRLSGPPGGVNSQGRKGVKSGSRGVPLP', 209-281, 'PV  
 15', 'D', '817-1260', 'P', '1262-1515' <GUI>  
 A:Cross-references: EMBL:X56979  
 R:Guo, X.; Kramer, J.M.  
 J. Biol. Chem. 264, 17574-17582, 1989  
 A:Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1  
 A:Reference number: A34476; MUID:90008929; PMID:2793871  
 A:Accession: B34476  
 A:Molecule type: DNA  
 A:Residues: 1432-1499, 'Q', '1501-1707', 'P', '1709-1744' <GU2>  
 A:Cross-references: EMBL:J05067; NID:g156255; PIDN:AAB59179.1; PID:g156256  
 C:Genetics:  
 A:Gene: clb-2; emb-9  
 A:Map position: 3  
 A:Inserts: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;  
 C:Superfamily: collagen alpha 1(IV) chain  
 C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
 F:43-1515/Domain: collagenous, triple helix #status predicted <COL>  
 F:93-95/Region: cell attachment (R-G-D) motif  
 F:1053-1055/Region: cell attachment (R-G-D) motif  
 F:1396-1398/Region: cell attachment (R-G-D) motif  
 F:1516-1744/Domain: carboxyl-terminal nonhelical, N1 #status predicted <NC1>  
 F:1516-1627, 1628-1744/Region: duplication  
 F:1580-1586, 1691-1697/Disulfide bonds: #status predicted

Query Match 68.2%; Score 75; DB 2; Length 1744;  
 Best Local Similarity 66.7%; Pred. No. 0.0057;  
 Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFSRNDYS 19  
 |||:|||||:|||||  
 Db 1578 MFCNMNSVCHVSRNDYS 1595

Search completed: April 5, 2004, 07:05:39  
 Job time : 4.91041 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 2.39225 Seconds  
(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFNCVNCVCFNSRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	93	84.5	1670	1 CA34 HUMAN	Q01955 homo sapien
2	92	83.6	471	1 CA34 BOVIN	Q28084 bos taurus
3	92	83.6	1669	1 CA14 HUMAN	P02462 homo sapien
4	92	83.6	1669	1 CA14 MOUSE	P02462 mus musculus
5	90	81.8	754	1 CA54 CANFA	Q28247 canis famil
6	90	81.8	1685	1 CA54 HUMAN	P29400 homo sapien
7	76	69.1	1763	1 CA24 ASCSU	P27393 ascaris suu
8	75	68.2	1758	1 CA14 CAEEL	P17139 caenorhabdi
9	67	60.9	1758	1 CA24 CAEEL	P17140 caenorhabdi
10	64	58.2	1691	1 CA64 HUMAN	Q14031 homo sapien
11	59	53.6	623	1 CA44 RABIT	P55787 oryctolagus
12	59	53.6	1690	1 CA44 HUMAN	P53420 homo sapien
13	59	53.6	1775	1 CA14 DROME	P08120 drosophila
14	58	52.7	453	1 CA44 BOVIN	Q29442 bos taurus
15	58	52.7	1707	1 CA24 MOUSE	P08122 mus musculus
16	58	52.7	1712	1 CA24 HUMAN	P08572 homo sapien
17	47	42.7	1477	1 HTK7 HYDAT	Q25197 hydra atten
18	44	40.0	333	1 AMR1 HUMAN	Q9y4x0 homo sapien
19	44	40.0	344	1 AMR1 MOUSE	Q9y4x0 mus musculus
20	42.5	38.6	553	1 VH65 NPVAC	Q08539 autographa
21	42.5	38.6	578	1 VAC8 YEAST	P39968 saccharomyc
22	42	38.2	288	1 SPY3 HUMAN	Q43610 homo sapien
23	42	38.2	312	1 XTC CALSA	P23557 caldocellum
24	42	38.2	339	1 OTN EUCBP	Q89891 buchnera ap
25	42	38.2	461	1 TR1B HUMAN	P20333 homo sapien
26	42	38.2	474	1 TR1B MOUSE	P25119 mus musculus
27	42	38.2	589	1 SPY DROME	Q44783 drosophila
28	42	38.2	703	1 NH55 CAEEL	Q9na51 caenorhabdi
29	42	38.2	1216	1 ATU1 YEAST	P38360 saccharomyc
30	42	38.2	6629	1 RIAB IBVSE	P27920 a replicase
31	42	38.2	6629	1 RIAB IBVSC	Q91qt2 a replicase
32	41	37.3	60	1 IT11 BRANA	P80301 brassica na
33	41	37.3	184	1 MTR2 YEAST	P34232 saccharomyc

RESULT 1  
CA34 HUMAN  
ID CA34 HUMAN STANDARD; PRT; 1670 AA.  
AC Q01955; O9BQT2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94364994; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain;  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RL J. Biol. Chem. 269:23013-23017(1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;  
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND  
RP CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;  
RP PRO-574; GLU-1269 AND PRO-1474.  
RX MEDLINE=21064696; PubMed=11134255;  
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
RN [4]  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:19780-19784(1992).  
RN [5]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=93353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554(1991).  
RN [6]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;

34 41 37.3 292 1 EFTS RALSO Q8xzj0 ralstonia s  
35 41 37.3 313 1 EFTS ANASP Q8vny3 anabaena sp  
36 41 37.3 327 1 CPZ7\_PIG P79492 sus scrofa  
37 41 37.3 365 1 H182 BRAJA Q89ul9 bradyrhizob  
38 41 37.3 480 1 CPC1 RABIT P00180 oryctolagus  
39 41 37.3 489 1 CPZ6 CANFA O62671 canis famil  
40 41 37.3 490 1 CPC2 RABIT P00181 oryctolagus  
41 41 37.3 490 1 CPCR MSAU P33264 mesocricetu  
42 41 37.3 506 1 Y619 METJA Q58036 methanococc  
43 41 37.3 567 1 PRR3 STRPU P50473 strongyloce  
44 41 37.3 623 1 PRR3 CANAL P46030 candida alb  
45 41 37.3 2476 1 ZAN\_PIG Q28983 sus scrofa







CC [3] SEQUENCE OF 1-943 FROM N.A.  
CC RP TISSUE=Placenta;  
CC RX MEDLINE=88029471; PubMed=3311751;  
CC RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,  
CC RA Deutzmann R., Kuehn K.;  
CC RT "Completion of the amino acid sequence of the alpha 1 chain of human  
CC RT basement membrane collagen (type IV) reveals 21 non-triplet  
CC RT interruptions located within the collagenous domain.";  
CC RL Eur. J. Biochem. 168:529-536(1987).  
CC RN [4]  
CC RN SEQUENCE OF 28-243.  
CC RP MEDLINE=86004708; PubMed=4043082;  
CC RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
CC RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
CC RT region (75 domain) of the alpha 1 (IV) chain of human basement  
CC RT membrane collagen.";  
CC RL Eur. J. Biochem. 152:213-219(1985).  
CC RN [5]  
CC RN SEQUENCE OF 534-1447.  
CC RP MEDLINE=8503629; PubMed=6434307;  
CC RA Babel W., Glanville R.W.;  
CC RT "Structure of human-basement-membrane (type IV) collagen. Complete  
CC RT amino-acid sequence of a 914-residue-long peptic fragment from the  
CC RT alpha 1(IV) chain.";  
CC RL Eur. J. Biochem. 143:545-556(1984).  
CC RN [6]  
CC RN SEQUENCE OF 1256-1669 FROM N.A.  
CC RP MEDLINE=85207819; PubMed=2581969;  
CC RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
CC RA Cheung M.-C., Prockop D.J., Boyd C.D.;  
CC RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV  
CC RT procollagen reveal an unusual homology of amino acid sequences in two  
CC RT halves of the carboxyl-terminal domain.";  
CC RL J. Biol. Chem. 260:7681-7687(1985).  
CC RN [7]  
CC RN SEQUENCE OF 1259-1669 FROM N.A.  
CC RP MEDLINE=85216555; PubMed=2582422;  
CC RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
CC RA Kefalides N.A., Myers J.C.;  
CC RT "Restricted homology between human alpha 1 type IV and other  
CC RT procollagen chains.";  
CC RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
CC RN [8]  
CC RN SEQUENCE OF 1-28 FROM N.A.  
CC RP MEDLINE=89034231; PubMed=3182844;  
CC RA Soiminen R., Ruotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
CC RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
CC RT collagen are divergently encoded on opposite DNA strands and have an  
CC RT overlapping promoter region.";  
CC RL J. Biol. Chem. 263:17217-17220(1988).  
CC RN [9]  
CC RN SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
CC RP TISSUE=Placenta;  
CC RX MEDLINE=89005112; PubMed=2844531;  
CC RA Siebold B., Deutzmann R., Kuehn K.;  
CC RT "The arrangement of intra- and intermolecular disulfide bonds in the  
CC RT carboxy-terminal, non-collagenous aggregation and cross-linking domain  
CC RT of basement-membrane type IV collagen.";  
CC RL Eur. J. Biochem. 176:617-624(1988).  
CC CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC CC meshwork together with laminins, proteoglycans and entactin/  
CC CC nidogen.  
CC CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC CC alpha 6(IV), each of which can form a triple helix structure  
CC CC with 2 other chains to generate type IV collagen network.  
CC CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC CC G-X-Y repeats in the long central triple-helical domain (which may  
CC CC cause flexibility in the triple helix), and a short N-terminal  
CC CC triple-helical 7S domain.  
CC CC -!- PTM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M26576; AAA53098.1; JOINED.  
CC EMBL; J04217; AAA53098.1; JOINED.  
CC EMBL; M26550; AAA53098.1; JOINED.  
CC EMBL; M26540; AAA53098.1; JOINED.  
CC EMBL; M26542; AAA53098.1; JOINED.  
CC EMBL; M26543; AAA53098.1; JOINED.  
CC EMBL; M26544; AAA53098.1; JOINED.  
CC EMBL; M26545; AAA53098.1; JOINED.  
CC EMBL; M26546; AAA53098.1; JOINED.  
CC EMBL; M26547; AAA53098.1; JOINED.  
CC EMBL; M26537; AAA53098.1; JOINED.  
CC EMBL; M26538; AAA53098.1; JOINED.  
CC EMBL; M26548; AAA53098.1; JOINED.  
CC EMBL; M26549; AAA53098.1; JOINED.  
CC EMBL; M26551; AAA53098.1; JOINED.  
CC EMBL; M26552; AAA53098.1; JOINED.  
CC EMBL; M26553; AAA53098.1; JOINED.  
CC EMBL; M26554; AAA53098.1; JOINED.  
CC EMBL; M26555; AAA53098.1; JOINED.  
CC EMBL; M26556; AAA53098.1; JOINED.  
CC EMBL; M26557; AAA53098.1; JOINED.  
CC EMBL; M26539; AAA53098.1; JOINED.  
CC EMBL; M26558; AAA53098.1; JOINED.  
CC EMBL; M26559; AAA53098.1; JOINED.  
CC EMBL; M26560; AAA53098.1; JOINED.  
CC EMBL; M26561; AAA53098.1; JOINED.  
CC EMBL; M26562; AAA53098.1; JOINED.  
CC EMBL; M26536; AAA53098.1; JOINED.  
CC EMBL; M26563; AAA53098.1; JOINED.  
CC EMBL; M26541; AAA53098.1; JOINED.  
CC EMBL; M26564; AAA53098.1; JOINED.  
CC EMBL; M26565; AAA53098.1; JOINED.  
CC EMBL; M26566; AAA53098.1; JOINED.  
CC EMBL; M26567; AAA53098.1; JOINED.  
CC EMBL; M26568; AAA53098.1; JOINED.  
CC EMBL; M26569; AAA53098.1; JOINED.  
CC EMBL; M26570; AAA53098.1; JOINED.  
CC EMBL; M26571; AAA53098.1; JOINED.  
CC EMBL; M26572; AAA53098.1; JOINED.  
CC EMBL; M26573; AAA53098.1; JOINED.  
CC EMBL; M26574; AAA53098.1; JOINED.  
CC EMBL; M26575; AAA53098.1; JOINED.  
CC EMBL; Y00706; CAA66698.1; -  
CC EMBL; X05561; CAA29075.1; -  
CC EMBL; M10940; AAA52006.1; -  
CC EMBL; M11315; AAA52042.1; -  
CC PTR; S16876; CGHU4B.  
CC Genew; HGNC:2202; COL4A1.  
CC MIW; 120130; -  
CC InterPro; IPR008161; C1g\_helix.  
CC InterPro; IPR008160; Collagen.  
CC InterPro; IPR001442; Procollagn4\_C.  
CC Pfam; PF01413; C4; 2.  
CC Pfam; PF01391; Collagen; 24.  
CC ProDom; PD000007; C1g\_helix; 6.  
CC ProDom; PD003923; ProcollagnC4; 1.  
CC SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Basement membrane;  
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1441 1669 NON-HELICAL REGION (NC1).  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).  
 FT DSULFID 1460 1551 OR 1548.  
 FT DSULFID 1493 1548 OR 1551.  
 FT DSULFID 1505 1511  
 FT DSULFID 1570 1665 OR 1662.  
 FT DSULFID 1604 1662 OR 1665.  
 FT DSULFID 1616 1622  
 FT CONFLICT 237 238 SG -> KE (IN REF. 4).  
 FT CONFLICT 241 241 G -> K (IN REF. 4).  
 FT CONFLICT 319 319 Q -> A (IN REF. 3).  
 FT CONFLICT 719 719 N -> D (IN REF. 5).  
 FT CONFLICT 837 837 D -> Y (IN REF. 5).  
 FT CONFLICT 842 842 K -> P (IN REF. 5).  
 FT CONFLICT 896 896 V -> W (IN REF. 2).  
 FT CONFLICT 904 904 E -> Q (IN REF. 5).  
 FT CONFLICT 914 914 S -> K (IN REF. 5).  
 FT CONFLICT 998 998 S -> K (IN REF. 5).  
 FT CONFLICT 1010 1010 K -> P (IN REF. 5).  
 FT CONFLICT 1012 1012 S -> K (IN REF. 5).  
 FT CONFLICT 1358 1358 E -> Q (IN REF. 5).  
 SQ SEQUENCE 1669 AA; 160611 MW; 3BEBAGDFFB9B8A84 CRC64;  
 Query Match 83.6%; Score 92; DB 1; Length 1669;  
 Best Local Similarity 88.9%; Pred. No. 3.4e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 LFCNVNVCNFCASRNNDYS 19  
 Db 1503 LFCNINNVCNFCASRNNDYS 1520  
 RESULT 4  
 CA14\_MOUSE STANDARD; PRT; 1669 AA.  
 AC P02463;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=8917932; PubMed=2703490;  
 RA Muthukumar G., Blumberg B., Kurkinen M.;  
 RT "The complete primary structure for the alpha 1-chain of mouse  
 RT collagen IV. Differential evolution of collagen IV domains.";  
 RL J. Biol. Chem. 264:6310-6317(1989).  
 RN [2]  
 RN SEQUENCE OF 1-1154 FROM N.A.  
 RX MEDLINE=89112221; PubMed=3338568;  
 RA Wood L., Theriault N., Vogeli G.;  
 RT "cDNA clones completing the nucleotide and derived amino acid  
 RT sequence of the alpha 1 chain of basement membrane (type IV) collagen  
 RT from mouse".  
 RL FEBS Lett. 227:5-8(1988).  
 RN [3]  
 RN SEQUENCE OF 1149-1424 FROM N.A.  
 RX MEDLINE=86301886; PubMed=3755692;  
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;  
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a  
 RT synthetic oligodeoxynucleotide".  
 RL Gene 43:301-304(1986).  
 RN [4]  
 RN SEQUENCE OF 1276-1669 FROM N.A.  
 RX MEDLINE=85127033; PubMed=2578961;  
 RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
 RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;  
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of  
 RT the alpha 1(IV) chain of basement membrane collagen as derived from  
 RT complementary DNA".  
 RL Eur. J. Biochem. 147:217-224(1985).  
 RN [5]  
 RN SEQUENCE OF 1441-1669 FROM N.A.  
 RX MEDLINE=87250460; PubMed=3597383;  
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
 RA Saus J., Pihlajaniemi T.;  
 RT "Extensive homology between the carboxyl-terminal peptides of mouse  
 RT alpha 1(IV) and alpha 2(IV) collagen.";  
 RL J. Biol. Chem. 262:8496-8499(1987).  
 RN [6]  
 RN PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=86196099; PubMed=3009468;  
 RA Sakurai Y., Sullivan M., Yamada Y.;  
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
 RT collagen genes".  
 RL J. Biol. Chem. 261:6654-6657(1986).  
 RN [7]  
 RN SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89066738; PubMed=3198626;  
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
 RT "Head-to-head arrangement of murine type IV collagen genes.";  
 RL J. Biol. Chem. 263:19274-19277(1988).  
 RN [8]  
 RN SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89071759; PubMed=3200851;  
 RA Burdello P.D., Martin G.R., Yamada Y.;  
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
 RT bidirectional promoter and a shared enhancer.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).  
 RN [9]  
 RN SEQUENCE OF 1-129 FROM N.A.  
 RX MEDLINE=88243724; PubMed=3379041;  
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;  
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
 RT collagen chain and the corresponding region of the gene.";  
 RL J. Biol. Chem. 263:8706-8709(1988).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
 CC alpha 6(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, J03758; AAA37439.1; -  
 DR EMBL, M23333; AAA51625.1; -

```
DR EMBL: J04694; AAA50292.1; -
DR EMBL: X06777; CAA29945.1; -
DR EMBL: X02201; CAA26132.1; -
DR EMBL: M15832; AAA37340.1; -
DR EMBL: M14042; AAA37342.1; -
DR EMBL: M12879; AAA37343.1; -
DR EMBL: M13024; -; NOT_ANNOTATED_CDS.
DR EMBL: M13025; -; NOT_ANNOTATED_CDS.
DR EMBL: M13026; AAA37344.1; -
DR EMBL: M13027; AAA37345.1; -
DR EMBL: M13043; AAA37346.1; -
DR EMBL: J04448; AAA37437.1; -
DR PIR: A33525; CGMS4B.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Clg helix; 6.
DR ProDom; PD001923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 BY SIMILARITY.
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 126 126 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91B52058E9 CRC64;

Query Match 83.6%; Score 92; DB 1; Length 1669;
Best Local Similarity 88.9%; Pred. No. 3.4e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASRNDYS 19
DQ 1503 LFCNINNVNCFASRNDYS 1520
|||||:|||||
|||||:|||||

RESULT 5
CA54_CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; Tissue=Kidney;
RA MEDLINE=34224868; PubMed=8171024;
RA Zheng K., Thorner P.S., Marrano P., Bauml R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of
CC canine X-linked hereditary nephritis (HN), a disease similar to
CC that in humans (also referred to as Alport syndrome) characterized
CC by progressive renal failure and neurosensory deafness.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U07888; AAB60258.1; -.
DR PIR; A55267; A55267.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 1
FT DOMAIN <1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 >754 NONHELICAL REGION (NC1).
FT DISULFID 552 643 OR 640 (BY SIMILARITY).
FT DISULFID 585 640 OR 643 (BY SIMILARITY).
FT DISULFID 597 603 BY SIMILARITY.
FT DISULFID 662 ? OR 754 (BY SIMILARITY).
FT DISULFID 696 754 BY SIMILARITY.
FT DISULFID 708 714 BY SIMILARITY.
FT NON_TER 754 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 81.8%; Score 90; DB 1; Length 754;
Best Local Similarity 83.3%; Pred. No. 3.1e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNPFASRNDYS 19
DQ 595 MFCNINNVNCFASRNDYS 612
|||||:|||||
|||||:|||||
```

RESULT 6  
 CAS4\_HUMAN  
 ID CA54\_HUMAN STANDARD; PRT; 1685 AA.  
 AC P29400; Q16006; Q16126;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DE 01-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 5(IV) chain precursor.  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165049; PubMed=8120014;  
 RA Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RL J. Biol. Chem. 269:6608-6614(1994).  
 RN [2]  
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92316923; PubMed=1352287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 RN [3]  
 RP SEQUENCE OF 85-1685 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90337990; PubMed=2380186;  
 RA Pihlajaniemi T., Pohjola E.R., Myers J.C.;  
 RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";  
 RL J. Biol. Chem. 265:13758-13766(1990).  
 RN [4]  
 RP SEQUENCE OF 924-1685 FROM N.A.  
 RX MEDLINE=91169491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 RN [5]  
 RP SEQUENCE OF 914-1685 FROM N.A.  
 RX MEDLINE=90160375; PubMed=1689491;  
 RA Hostikka S.L., Eddy R.L., Syers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;  
 RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6]  
 RP SEQUENCE OF 1442-1471 FROM N.A.  
 RX MEDLINE=90252791; PubMed=2339699;  
 RA Myers J.C., Jones T.A., Pohjola E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;  
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";  
 RL Am. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).  
 RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;  
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [10]  
 RP VARIANT AS SER-1564.  
 RX MEDLINE=91169492; PubMed=1672282;  
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;  
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]  
 RP VARIANT AS ARG-325.  
 RX MEDLINE=92303559; PubMed=1376965;  
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";  
 RL Am. J. Hum. Genet. 51:135-142(1992).  
 RN [12]  
 RP VARIANT AS GLU-325.  
 RX MEDLINE=93244772; PubMed=1363780;  
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";  
 RL Hum. Mol. Genet. 1:127-129(1992).  
 RN [13]  
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
 RX MEDLINE=94010948; PubMed=8406498;  
 RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggema-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796; ARG-869; ARG-872 AND CYS-1241.  
 RX MEDLINE=95322976; PubMed=7599631;  
 RA Boye S., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the COL4A5 gene in Alport syndrome patients.";  
 RL Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=96213750; PubMed=8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M., Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is common in the western United States.";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS.  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Brutini M., Galli L., Zanelli P., Neri T.M., Rossetti S., Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G., Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F., Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51 exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).





```
FT isoform II).
FT /FTID=VSP 001159,
SQ SEQUENCE 1763 AA, 168526 MW, 304F528C06A80D CRC64;
Query Match 69.1%; Score 76; DB 1; Length 1763;
Best Local Similarity 77.8%; Pred. No. 0.00094;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 LFCNVNVCNFAASNDYS 19
DB 1591 LFCNVNVCNFAASNDYS 1608
RESULT 8
CA14 CAEL STANDARD; PRT; 1758 AA.
AC P17139;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN EMB-9 OR CLB-2 OR K04H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Guo X., Johnson J.J., Kramer J.M.;
RT "Embryonic lethality caused by mutations in basement membrane
RL collagen of C. elegans.";
RL Nature 349:707-709(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Karshaw J., Kitsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1446-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RL genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NC1
CC domains.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
```



OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Eye, and Kidney;  
RX MEDLINE=94171779; PubMed=8125972;  
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;  
RT Identification of a new collagen IV chain, alpha 6(IV), by cDNA  
RT isolation and assignment of the gene to chromosome Xq22, which is the  
RT same locus for COL4A5.";  
RL J. Biol. Chem. 269:7520-7526 (1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=94230418; PubMed=8175748;  
RA Zhou J., Ding M., Zhao Z., Reders S.T.;  
RT "Complete primary structure of the sixth chain of human basement  
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)  
RT and comparison with five other type IV collagen chains.";  
RL J. Biol. Chem. 269:13193-13199 (1994).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND  
RP LYS-1110.  
RX MEDLINE=96299642; PubMed=8661006;  
RA Zhang X., Zhou J., Reders S.T., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479 (1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Bird C., Grafham D., Lawlor S., Wilson S.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=93361972; PubMed=8356449;  
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,  
RA de Paeppe A., Tryggvason K., Reders S.F.;  
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RT inherited smooth muscle tumors.";  
RL Science 261:1167-1169 (1993).  
CC !- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC !- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC !- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC !- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequences=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequences=VSP 001174;  
CC !- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC !- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC !- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC !- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D21337; BAA04809.1; -;  
CC EMBL; U04845; AAA19569.2; -;  
CC EMBL; U47004; AAB19038.1; -;  
CC EMBL; U46959; AAB19038.1; JOINED.  
CC EMBL; U46961; AAB19038.1; JOINED.  
CC EMBL; U46962; AAB19038.1; JOINED.  
CC EMBL; U46963; AAB19038.1; JOINED.  
CC EMBL; U46964; AAB19038.1; JOINED.  
CC EMBL; U46965; AAB19038.1; JOINED.  
CC EMBL; U46966; AAB19038.1; JOINED.  
CC EMBL; U46967; AAB19038.1; JOINED.  
CC EMBL; U46968; AAB19038.1; JOINED.  
CC EMBL; U46969; AAB19038.1; JOINED.  
CC EMBL; U46970; AAB19038.1; JOINED.  
CC EMBL; U46971; AAB19038.1; JOINED.  
CC EMBL; U46972; AAB19038.1; JOINED.  
CC EMBL; U46973; AAB19038.1; JOINED.  
CC EMBL; U46974; AAB19038.1; JOINED.  
CC EMBL; U46975; AAB19038.1; JOINED.  
CC EMBL; U46976; AAB19038.1; JOINED.  
CC EMBL; U46977; AAB19038.1; JOINED.  
CC EMBL; U46978; AAB19038.1; JOINED.  
CC EMBL; U46979; AAB19038.1; JOINED.  
CC EMBL; U46980; AAB19038.1; JOINED.  
CC EMBL; U46981; AAB19038.1; JOINED.  
CC EMBL; U46982; AAB19038.1; JOINED.  
CC EMBL; U46983; AAB19038.1; JOINED.  
CC EMBL; U46984; AAB19038.1; JOINED.  
CC EMBL; U46985; AAB19038.1; JOINED.  
CC EMBL; U46986; AAB19038.1; JOINED.  
CC EMBL; U46987; AAB19038.1; JOINED.  
CC EMBL; U46988; AAB19038.1; JOINED.  
CC EMBL; U46989; AAB19038.1; JOINED.  
CC EMBL; U46990; AAB19038.1; JOINED.  
CC EMBL; U46991; AAB19038.1; JOINED.  
CC EMBL; U46992; AAB19038.1; JOINED.  
CC EMBL; U46993; AAB19038.1; JOINED.  
CC EMBL; U46994; AAB19038.1; JOINED.  
CC EMBL; U46995; AAB19038.1; JOINED.  
CC EMBL; U46996; AAB19038.1; JOINED.  
CC EMBL; U46997; AAB19038.1; JOINED.  
CC EMBL; U46998; AAB19038.1; JOINED.  
CC EMBL; U46999; AAB19038.1; JOINED.  
CC EMBL; U47000; AAB19038.1; JOINED.  
CC EMBL; U47001; AAB19038.1; JOINED.  
CC EMBL; U47002; AAB19038.1; JOINED.  
CC EMBL; U47003; AAB19038.1; JOINED.  
CC EMBL; U47004; AAB19038.1; JOINED.  
CC EMBL; U46960; AAB19039.1; JOINED.  
CC EMBL; U46961; AAB19039.1; JOINED.  
CC EMBL; U46962; AAB19039.1; JOINED.  
CC EMBL; U46963; AAB19039.1; JOINED.  
CC EMBL; U46964; AAB19039.1; JOINED.  
CC EMBL; U46965; AAB19039.1; JOINED.  
CC EMBL; U46966; AAB19039.1; JOINED.  
CC EMBL; U46967; AAB19039.1; JOINED.  
CC EMBL; U46968; AAB19039.1; JOINED.  
CC EMBL; U46969; AAB19039.1; JOINED.  
CC EMBL; U46970; AAB19039.1; JOINED.  
CC EMBL; U46971; AAB19039.1; JOINED.  
CC EMBL; U46972; AAB19039.1; JOINED.  
CC EMBL; U46973; AAB19039.1; JOINED.  
CC EMBL; U46974; AAB19039.1; JOINED.  
CC EMBL; U46975; AAB19039.1; JOINED.  
CC EMBL; U46976; AAB19039.1; JOINED.  
CC EMBL; U46977; AAB19039.1; JOINED.  
CC EMBL; U46978; AAB19039.1; JOINED.  
CC EMBL; U46979; AAB19039.1; JOINED.  
CC EMBL; U46980; AAB19039.1; JOINED.  
CC EMBL; U46981; AAB19039.1; JOINED.  
CC EMBL; U46982; AAB19039.1; JOINED.  
CC EMBL; U46983; AAB19039.1; JOINED.

RESULT 12
CA44_HUMAN
ID CA44
AC P5341
DT 01-O
DT 01-O
DT 28-F
DE Colla
GN COL4
OS Homo

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=95014445; PubMed=7523402;  
RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reiders S.T.;  
RT "Complete primary structure of the human type IV collagen alpha 4 (IV)  
RT chain. Comparison with structure and expression of the other alpha  
RT (IV) chains.";  
RL J. Biol. Chem. 269:26172-26177(1994).  
RN [2]  
RP SEQUENCE OF 1-23 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Ninomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q36.";  
RL FEBS Lett. 424:11-16(1998).  
RN [3]  
RP SEQUENCE OF 1219-1690 FROM N.A.  
RC TISSUE=Eye;  
RX MEDLINE=93374047; PubMed=8365481;  
RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;  
RT "cDNA isolation and partial gene structure of the human alpha 4(IV)  
RT collagen chain.";  
RL FEBS Lett. 330:122-128(1993).  
RN [4]  
RP SEQUENCE OF 1407-1507 FROM N.A.  
RX MEDLINE=93054733; PubMed=1429714;  
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;  
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
RT alpha 4 chain of basement membrane collagen type IV and assignment of  
RT the gene to the distal long arm of human chromosome 2.";  
RL J. Biol. Chem. 267:23753-23758(1992).  
RN [5]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97338662; PubMed=9195222;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
RL Hum. Mutat. 9:477-499(1997).  
RN [6]  
RP VARIANT AS SER-1201.  
RX MEDLINE=95078927; PubMed=7987396;  
RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
RA Smeets H.J.M., Reiders S.T.;  
RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)  
RT collagen genes in autosomal recessive Alport syndrome.";  
RL Nat. Genet. 8:77-82(1994).  
RN [7]  
RP VARIANT FBH GLU-897.  
RX MEDLINE=96379660; PubMed=8787673;  
RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
RT "Benign familial hematuria due to mutation of the type IV collagen  
RT alpha4 gene.";  
RL J. Clin. Invest. 98:1114-1118(1996).  
RN [8]  
RP VARIANTS AS AND VARIANTS.  
RX MEDLINE=99011253; PubMed=9792860;  
RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,  
RA Cochat P., Gruenfeld J.-P., Falcoix J.-B., Gubler M.-C., Antignac C.;  
RT "Determination of the genomic structure of the COL4A4 gene and of  
RT novel mutations causing autosomal recessive Alport syndrome.";  
RL Am. J. Hum. Genet. 63:1329-1340(1998).  
CC -1- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
alpha 6(IV), each of which can form a triple helix structure with  
2 other chains to generate type IV collagen network.  
-1- SUBCELLULAR LOCATION: Cell surface (Potential).  
-1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
colocalized and present only in basement membranes of kidney, eye,  
cochlea, lung and brain.  
-1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the G-  
X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
-1- PTM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
-1- PTM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
-1- DISEASE: Defects in COL4A4 are a cause of autosomal recessive  
Alport syndrome (AS) [MIM:203780], an hereditary disorder  
characterized by progressive glomerulonephritis, renal failure,  
hematuria, ocular abnormalities and deafness. The recessive form  
occurs equally between males and females.  
-1- DISEASE: Defects in COL4A4 are a cause of familial benign  
hematuria (FBH) [MIM:141200] or thin basement membrane disease.  
FBH is characterized by persistent hematuria, an electron  
microscopically detectable thin glomerular basement membrane (GBM)  
and an autosomal dominant mode of inheritance. Renal function  
remains normal. In children, differentiation between FBH and AS  
can be difficult, because both disorders are manifested by  
persistent hematuria and thin GBM at that age.  
-1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL; X81053; CAA56943.1; --  
EMBL; AB008496; BAA25065.1; --  
EMBL; D17391; BAA04214.1; --  
PIR; A55360; CGHUIB.  
Gene; HGNC:2206; COL4A4.  
MIM; 120131; --  
MIM; 141200; --  
MIM; 203780; --  
InterPro; IPR008161; C1g\_helix.  
InterPro; IPR008160; Collagen.  
InterPro; IPR001442; ProcollagN4\_C.  
Pfam; PF01413; C4; 2.  
Pfam; PF01391; Collagen; 21.  
ProDom; PD000007; C1g\_helix; 3.  
ProDom; PD003923; ProcollagN4; 1.  
SMART; SM00111; C4; 2.  
Extracellular matrix; Connective tissue; Basement membrane; Repeat;  
Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;  
KW Polymorphism; Alport syndrome.  
FT SIGNAL 1 38  
FT CHAIN 39 1690  
FT DOMAIN 39 64  
FT DOMAIN 65 1459  
FT DOMAIN 1460 1690  
FT SITE 94 96  
FT SITE 145 147  
FT SITE 189 191  
FT SITE 310 312  
FT SITE 724 726  
FT SITE 785 787  
FT SITE 989 991  
FT SITE 1206 1207  
FT COLLAGEN ALPHA 4(IV) CHAIN.  
FT 7S DOMAIN.  
FT TRIPLE-HELICAL REGION.  
FT NONHELICAL REGION (NC1).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT CLEAVAGE (BY COLLAGENASE)  
FT (BY SIMILARITY).



```

FT DISULFID 1720 1727 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC... ) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 ETGNV -> RAGR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AB23149525CD CRC64;

Query Match 53.6%; Score 59; DB 1; Length 1775;
Best Local Similarity 68.8%; Pred. No. 0.36;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRND 17
| | | | |
Db 1609 LSCGQNNVCNYSRND 1624

RESULT 14
CA44_BOVIN STANDARD; PRT; 453 AA.
AC Q29442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 4 (IV) chain (fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Redders S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens.";
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]
RP SEQUENCE OF 217-246.
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 217-233.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: M23334; AAA51626.1; -.

[illegible]

```

Query Match          52.7%  Score 58; DB 1; Length 1707;
Best Local Similarity 61.1%  Pred. NO. 0.49;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      2 LFCNVNVCVNPASRNDYS 19
      | : | | | | |
Db      1542 LYCNPGDVCCYASRNDXS 1559

Search completed: April 5, 2004, 06:59:42
Job time : 4.39225 secs

```

MEDLINE=89066738; PubMed=31398626;  
Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
"Head-to-head arrangement of murine type IV collagen genes.";  
J. Biol. Chem. 263:19274-19277(1988).  
[3]  
SEQUENCE OF 970-1480 FROM N.A.  
MEDLINE=86220192; PubMed=3011432;  
Schwarz U., Schuppan D., Oberbaumer I., Glanville R.W.,  
Deutzmann R., Timpi R., Kuehn K.;  
"Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain.";  
Eur. J. Biochem. 157:49-56(1986).  
[4]  
SEQUENCE OF 1480-1707 FROM N.A.  
MEDLINE=87054581; PubMed=3780963;  
Schwarz-Magdolen I., Oberbaumer I., Kuehn K.;  
"CDNA and protein sequence of the NC1 domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV).";  
FEBS Lett. 208:203-207(1986).  
[5]  
SEQUENCE OF 1481-1707 FROM N.A.  
MEDLINE=87250460; PubMed=3597383;  
Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S., Saus J., Pihlajaniemi T.;  
"Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";  
J. Biol. Chem. 262:8496-8499(1987).  
[6]  
SEQUENCE OF 1041-1489 FROM N.A.  
MEDLINE=87005245; PubMed=3758345;  
Vogeli G., Horn E., Carter J., Kaytes P.S.;  
"Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";  
FEBS Lett. 206:29-32(1986).  
[7]  
SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.  
MEDLINE=85296379; PubMed=3839908;  
Kurkinen M., Barnard M.P., Barlow D.P., Chow L.T.;  
"Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";  
Nature 317:177-179(1985).  
[8]  
SEQUENCE OF 1-60 FROM N.A.  
MEDLINE=89071759; PubMed=3200851;  
Burbelo P.D., Martin G.R., Yamada Y.;  
"Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer";  
Proc. Natl. Acad. Sci. U.S.A. 85:9679-9683(1988).  
-I- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and actinin/nidogen.  
CC  
-I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.  
CC  
-I- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.  
CC  
-I- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC  
-I- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.  
CC  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial purposes is prohibited.  
CC

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 11.4092 Seconds  
(without alignments)  
525.440 Million cell updates/sec

Title: US-10-032-221b-41

Perfect score: 110

Sequence: 1 KLFNCNVNVCNPNASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriopl:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	84.5	212	6 Q28512	Q28512 macaca mula
2	93	84.5	245	4 Q9NYC4	Q9NYC4 homo sapien
3	92	83.6	161	11 Q61430	Q61430 mus musculus
4	92	83.6	203	6 Q29032	Q29032 sus scrofa
5	92	83.6	203	6 Q28682	Q28682 cryctolagus
6	92	83.6	210	6 Q28273	Q28273 canis famil
7	92	83.6	212	6 Q28567	Q28567 ovis aries
8	92	83.6	225	6 Q28271	Q28271 canis famil
9	92	83.6	226	11 Q93LQ8	Q93LQ8 mus musculus
10	92	83.6	229	4 Q8NF88	Q8NF88 homo sapien
11	92	83.6	229	4 Q9NYC5	Q9NYC5 homo sapien
12	92	83.6	246	11 Q61435	Q61435 mus musculus
13	92	83.6	979	13 Q919K3	Q919K3 gallus gall
14	92	83.6	1075	4 Q86X41	Q86X41 homo sapien
15	92	83.6	1621	4 Q9H4R9	Q9H4R9 homo sapien
16	92	83.6	1669	11 Q9QZS0	Q9QZS0 mus musculus

17	90	81.8	179	11	P70165	P70165 mus musculus
18	90	81.8	253	11	Q61436	Q61436 mus musculus
19	90	81.8	585	11	Q80V57	Q80V57 mus musculus
20	90	81.8	799	11	Q8ENS7	Q8ENS7 mus musculus
21	90	81.8	886	4	Q9NUB7	Q9NUB7 homo sapien
22	90	81.8	1684	6	Q8HYC1	Q8HYC1 canis famil
23	90	81.8	1688	6	Q866Z2	Q866Z2 canis famil
24	90	81.8	1691	11	Q9ESQ2	Q9ESQ2 mus musculus
25	88	80.0	220	11	Q63122	Q63122 rattus norv
26	80	72.7	1752	5	Q07265	Q07265 strongyloce
27	76	69.1	1747	5	Q26640	Q26640 strongyloce
28	69	62.7	1802	5	Q17163	Q17163 brugia mala
29	64	58.2	205	6	Q28274	Q28274 canis famil
30	64	58.2	546	11	Q99K97	Q99K97 mus musculus
31	64	58.2	1600	4	Q9UEH6	Q9UEH6 homo sapien
32	64	58.2	1691	11	Q9ESQ1	Q9ESQ1 mus musculus
33	59	53.6	332	11	Q64457	Q64457 mus musculus
34	59	53.6	1024	5	Q8T7S4	Q8T7S4 anopheles g
35	59	53.6	1682	11	Q9QZR9	Q9QZR9 mus musculus
36	59	53.6	1723	5	Q9GQB1	Q9GQB1 hydra atten
37	59	53.6	1779	5	Q9VMV4	Q9VMV4 drosophila
38	58	52.7	202	6	Q28272	Q28272 canis famil
39	58	52.7	208	6	Q29468	Q29468 canis famil
40	58	52.7	358	11	Q91V13	Q91V13 mus musculus
41	58	52.7	673	4	Q14052	Q14052 homo sapien
42	54	49.1	713	5	Q9GV24	Q9GV24 sarcophaga
43	53	48.2	310	11	Q8JZZ6	Q8JZZ6 mus musculus
44	51	46.4	2275	3	Q93937	Q93937 emericella
45	49	44.5	346	4	Q9NS55	Q9NS55 homo sapien

#### ALIGNMENTS

#### RESULT 1

ID Q28512 PRELIMINARY; PRT; 212 AA.  
AC Q28512;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheidae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney cortex;  
RA Turner A.N., Ryan C.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Fusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different mammals";  
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
EL EMBL; L47280; AAA93861.1;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; ProcollagN4 C.  
DR InterPro; IPR000504; RNA\_rec\_mot-  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagN4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PRO00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1  
FT NON\_TER 212  
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357B64 CRC64;

Query Match 84.5%; Score 93; DB 6; Length 212;  
Best Local Similarity 94.4%; Pred. No. 1.7e-07;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
Db 45 LFCNVNVCNCFASRNDYS 62

## RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.  
ID Q9NYC4  
AC Q9NYC4  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Tmsctatin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hoffer H., Xiao Y., Stillman I.E., Kalluri R.,  
RT "Distinct anti-tumor properties of a type IV collagen domain derived  
RT from basement membrane."  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL: AF258351; AAF72632.1; -.  
DR GO: GO:0005581; C:collagen; IEA.  
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro: IPR001442; Procollagn4.C.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; ProcollagnC4; 1.  
DR SMART: SM00111; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; 1.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 84.5%; Score 93; DB 4; Length 245;  
Best Local Similarity 94.4%; Pred. No. 1.9e-07;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
Db 78 LFCNVNVCNCFASRNDYS 95

## RESULT 3

Q61430 PRELIMINARY; PRT; 161 AA.  
ID Q61430  
AC Q61430  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha3 chain (Fragment).  
GN COL4A3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Oberbaumer I.;  
RA STRAIN=129;  
RT "Cloning of the NC1 domains fo the minor collagen IV chains of mouse  
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and  
RT alpha5 (IV) in differentiated teratocarcinoma cells."  
RL Submitted (Oct-1994) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: X82205; CAA57689.1; -.  
DR PIR: S49488; S49488.  
DR GO: GO:0005581; C:collagen; IEA.  
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro: IPR001442; Procollagn4.C.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF01413; C4; 2.  
DR SMART: SMO0111; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; 1.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFB9236C5 CRC64;

Query Match 83.6%; Score 92; DB 11; Length 161;  
Best Local Similarity 88.9%; Pred. No. 1.9e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
Db 12 LFCNVNVCNCFASRNDYS 29

## RESULT 4

Q29032 PRELIMINARY; PRT; 203 AA.  
ID Q29032  
AC Q29032;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
RT mammals."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: L47284; AAA91882.1; -.  
DR GO: GO:0005581; C:collagen; IEA.  
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro: IPR001442; Procollagn4.C.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; ProcollagnC4; 1.  
DR SMART: SMO0111; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 83.6%; Score 92; DB 6; Length 203;  
Best Local Similarity 88.9%; Pred. No. 2.3e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
Db 45 LFCNVNVCNCFASRNDYS 62

## RESULT 5

Q28682 PRELIMINARY; PRT; 203 AA.  
ID Q28682  
AC Q28682;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.

```

OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47283; AAA91893.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 83.6%; Score 92; DB 6; Length 203;
Best Local Similarity 88.9%; Pred. NO. 2.3e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 45 LFCNINDVCNCFASRNDYS 62

RESULT 6
Q28273 ID Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119B4CA823633D CRC64;

Query Match 83.6%; Score 92; DB 6; Length 210;

```

```

Best Local Similarity 88.9%; Pred. NO. 2.4e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 55 LFCNINNVCNCFASRNDYS 72

RESULT 7
Q28567 ID Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 83.6%; Score 92; DB 6; Length 212;
Best Local Similarity 88.9%; Pred. NO. 2.4e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19
Db 45 LFCNINDVCNCFASRNDYS 62

RESULT 8
Q28271 ID Q28271 PRELIMINARY; PRT; 225 AA.
AC Q28271;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 1 chain (Fragment).
GN COL4A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains

```

RT of collagen type IV. Evidence from a canine model of X-linked  
 RT nephritis with a COL4A5 gene mutation.";  
 RL J. Biol. Chem. 271:13821-13828(1996).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-Saved;  
 RA Thorner P.S.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U50933; AAC48583.2; -;  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 FT NON\_TER 1  
 FT NON\_TER 225  
 SQ SEQUENCE 225 AA; 24585 MW; 2C20455890416E47 CRC64;

Query Match 83.6%; Score 92; DB 6; Length 225;  
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19  
 |||||:|||||  
 Db 69 LFCNNVNCVCFASRNDYS 86

RESULT 9  
 Q99LQ8 PRELIMINARY; PRT; 226 AA.  
 ID Q99LQ8  
 AC Q99LQ8  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN COL4A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC002269; AAH02269.1; -;  
 DR MGD; MGI:88454; Col4a1.  
 DR GO; GO:0005604; C:basement membrane; IEA.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 226 AA; 25042 MW; 4F7F0D5371181C21 CRC64;

Query Match 83.6%; Score 92; DB 11; Length 226;  
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19  
 |||||:|||||  
 Db 60 LFCNNVNCVCFASRNDYS 77

RESULT 10  
 Q8NF88 PRELIMINARY; PRT; 229 AA.  
 ID Q8NF88  
 AC Q8NF88  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Arresten (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA He A.B.;  
 RT "Cloning and Expression of Arresten in Escherichia coli and Pachia  
 RT pastoris.";  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF36207; AM97359.1; -;  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 229 AA; 25391 MW; 09B21FD5AB517E9E CRC64;  
 SQ SEQUENCE

Query Match 83.6%; Score 92; DB 4; Length 229;  
 Best Local Similarity 88.9%; Pred. No. 2.6e-07;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNNVNCVCFASRNDYS 19  
 |||||:|||||  
 Db 63 LFCNNVNCVCFASRNDYS 80

RESULT 11  
 Q9NYC5 PRELIMINARY; PRT; 229 AA.  
 ID Q9NYC5  
 AC Q9NYC5  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Arresten (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Colorado P.C.; Torre A.; Kamphaus G.D.; Maeshima Y.; Hopfer H.;  
 RA Takahashi K.; Volk R.; Zamborsky E.D.; Herman S.; Sarkar P.K.;  
 RA Ericken M.B.; Dhanabal M.; Simons M.; Post M.; Kufe D.;  
 RA Weichselbaum R.R.; Sukhatme V.P.; Kalluri R.;  
 RT "Anti-angiogenic cues from vascular basement membrane collagen.";  
 RL Cancer Res. 0:0-0(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Fu J.; Bai X.; Wang W.; Ruan C.;  
 RT "Arresten, a collagen-derived inhibitor of angiogenesis.";  
 RL Chung Hua Hsueh Yen Hsueh Tea Chih 22:0-0(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Peng X.; Yin B.; Yuan J.; Qiang B.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Zheng Q.C.; Song Z.F.; Zheng Y.W.; Li Y.Q.; Shu X.;  
 RT "Molecular cloning and sequencing of human arresten gene.";  
 RL Zhonghua Shi Yan Wei Ke Za Zhi 19:46-47(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Song Z.F.; Zheng Q.C.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF258349; AAF72630.1; -;  
 DR EMBL; AF363672; AAK53382.1; -;  
 DR EMBL; AF400431; AAK92480.1; -;  
 DR EMBL; AY285780; AAP43112.1; -;  
 DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
FT NON TER 1  
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;

Query Match 83.6%; Score 92; DB 4; Length 229;  
Best Local Similarity 88.9%; Pred. No. 2.6e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
DB 53 LFCNINNVNFCASRNDYS 80  
|||||:|||||

RESULT 12  
Q61435 PRELIMINARY; PRT; 246 AA.  
AC Q61435;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen IV alpha 3 chain (Fragment).  
GN COL4A3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/c;  
RX MEDLINE=95050957; PubMed=7962065;  
RA Miner J.H.; Sanes J.R.;  
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal  
laminae: Sequence, distribution, association with laminins, and  
developmental switches."  
RL J. Cell Biol. 127:879-891(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/c;  
RA Miner J.H.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z35166; CAA84529.1; -.  
DR PIR; I48302; I48302.  
DR MGD; MGI:104688; Col4a3.  
DR GO; GO:0005604; C:basement membrane; IDA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
FT NON TER 1  
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F58367324 CRC64;

Query Match 83.6%; Score 92; DB 11; Length 246;  
Best Local Similarity 88.9%; Pred. No. 2.8e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
DB 79 LFCNINNVNFCASRNDYS 96  
|||||:|||||

RESULT 13  
Q919K3 PRELIMINARY; PRT; 979 AA.  
AC Q919K3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Collagen IV al chain (Fragment).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Halfter W.M.; Dong S.;  
RT "Composition, synthesis and assembly of the embryonic chick retinal  
basal lamina."  
RL Dev. Biol. 0:0-0(2000).  
DR EMBL; AF239838; AAF44681.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD000007; C1g\_helix; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON TER 1  
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 83.6%; Score 92; DB 13; Length 979;  
Best Local Similarity 88.9%; Pred. No. 9.8e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
DB 813 LFCNINNVNFCASRNDYS 830  
|||||:|||||

RESULT 14  
Q86X41 PRELIMINARY; PRT; 1075 AA.  
AC Q86X41;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to collagen, type IV, alpha 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC047305; AAH47305.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD000007; C1g\_helix; 3.  
DR ProDom; PD003923; ProcollagnC4; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON TER 1  
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 83.6%; Score 92; DB 4; Length 1075;  
Best Local Similarity 88.9%; Pred. No. 1.1e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19

```

Db      909 LFCNNVCNCFASRNDYS 926
||||:| |||||
RESULT 15
Q9H4R9
ID Q9H4R9 PRELIMINARY; PRT; 1621 AA.
AC Q9H4R9
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1) (Fragment).
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390755; CAC13153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g_helix; 5.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDOEDBA2 CRC64;
Query Match 83.6%; Score 92; DB 4; Length 1621;
Best Local Similarity 88.9%; Pred. No. 1.6e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 LFCNNVCNCFASRNDYS 19
||||:| |||||
Db      1455 LFCNNVCNCFASRNDYS 1472

```

Search completed: April 5, 2004, 07:03:58  
 Job time : 11.4092 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time: 17.1138 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221b-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNPFASRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	19	6	ADA20240 TP3 peptid
2	93	84.5	25	6	ADA20236 T7 peptid
3	93	84.5	27	6	ADA20238 T8 peptid
4	93	84.5	79	5	Aau75600 Human typ
5	93	84.5	79	6	ADA20264 Human typ
6	93	84.5	88	5	Aau75608 Human typ
7	93	84.5	88	5	Aau75607 Human typ
8	93	84.5	88	6	ADA20271 Human typ
9	93	84.5	88	6	ADA20272 Human typ
10	93	84.5	124	5	Aau75594 Human typ
11	93	84.5	124	6	ADA20258 Human typ
12	93	84.5	132	5	Aau75597 Human typ
13	93	84.5	132	6	ADA20261 Human typ
14	93	84.5	131	5	Aau75596 Human typ
15	93	84.5	191	6	ADA20260 Human typ
16	93	84.5	211	3	Aay95918 Human Goo
17	93	84.5	211	5	ABG79208 Human GP
18	93	84.5	218	2	AAR79164 Partial s
19	93	84.5	218	2	Aay44172 Human typ
20	93	84.5	218	3	Aay56784 Human alp
21	93	84.5	218	4	Aae09484 Human alp
22	93	84.5	232	7	ADC17697 Human typ
23	93	84.5	244	5	ABG79218 Human typ
24	93	84.5	244	5	ABG79219 Human Goo
25	93	84.5	244	5	ABG79217 Human typ

26	93	84.5	244	5	Aau75595 Human typ
27	93	84.5	244	6	ADA20235 Human typ
28	93	84.5	245	3	Aay67942 Human typ
29	93	84.5	245	5	Aau75589 Human typ
30	93	84.5	254	5	Aau75598 Human typ
31	93	84.5	268	2	Aay31993 Type IV c
32	93	84.5	268	3	Aay97555 Human alp
33	93	84.5	1670	7	ADA47063 Human Pro
34	92	83.6	229	1	AAP93524 Complete
35	92	83.6	229	3	Aay67943 Human typ
36	92	83.6	229	5	Aau75587 Human typ
37	92	83.6	229	6	ADA20217 Human typ
38	92	83.6	229	7	ADC17695 Human typ
39	92	83.6	260	2	Aay31991 Type IV c
40	92	83.6	260	3	Aay97553 Human alp
41	92	83.6	406	3	AAB58169 Lung canc
42	92	83.6	471	2	AAR79163 Partial s
43	92	83.6	471	2	Aay44171 Bovine ty
44	92	83.6	471	3	Aay56783 Bovine al
45	92	83.6	471	4	Aae09483 Bovine al

## ALIGNMENTS

RESULT 1

ADA20240  
ID ADA20240 standard; peptide; 19 AA.

XX AC ADA20240;

XX DT 20-NOV-2003 (first entry)

XX DE TP3 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytostatic; gene therapy; TP3 peptide; tumstatin; human;  
KW type IV collagen alpha 3 chain; mutant; mutein.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Phe substituted by Lys"

FT Misc-difference 8 /note= "Wild-type Asp substituted by Cys"

XX WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 64; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the TP3 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.

XX SQ Sequence 19 AA;

Query Match 100.0%; Score 110; DB 6; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLFCNVNVCNCFASRNDYS 19  
 |||||  
 DB 1 KLFCNVNVCNCFASRNDYS 19

# RESULT 2

ADA20236  
 ID ADA20236 standard; peptide; 25 AA.

XX AC ADA20236;

XX DT 20-NOV-2003 (first entry)

XX DE T7 peptide related to human type IV collagen alpha and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX KW metastasis; basement membrane organisation; type IV collagen network;  
 XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX KW cytostatic; gene therapy; T7 peptide; tumstatin; human;  
 XX KW type IV collagen alpha 3 chain.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor  
 XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 53; Page 45; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the T7 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.

XX SQ Sequence 25 AA;

Query Match 84.5%; Score 93; DB 6; Length 25;  
 Best Local Similarity 94.4%; Pred. No. 5.8e-07;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19  
 |||||

DB 5 LFCNVNVCNCFASRNDYS 22

# RESULT 3

ADA20238  
 ID ADA20238 standard; peptide; 27 AA.

XX AC ADA20238;

XX DT 20-NOV-2003 (first entry)

XX DE T8 peptide related to human type IV collagen alpha and angiogenesis.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX KW metastasis; basement membrane organisation; type IV collagen network;  
 XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX KW cytostatic; gene therapy; T8 peptide; tumstatin; human;  
 XX KW type IV collagen alpha 3 chain; mutant; mutein.

XX OS Synthetic.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

XX FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

XX FT WO2003059257-A2.

XX PN 24-JUL-2003.

XX PD 20-DEC-2002; 2002WO-US040938.

XX PF 21-DEC-2001; 2001US-00032221.

XX PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX PT New peptide, useful for preparing a composition for inhibiting tumor  
 XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 62; Page 45; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.



XX Kalluri R;  
 XX WPI; 2003-587256/55.  
 DR N-PSDB; ADA20224.  
 XX  
 XX New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 PT  
 XX Claim 94; SEQ ID NO 26; 240pp; English.  
 PS  
 XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumor growth and metastasis.  
 CC Basement membrane organization is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumor  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.  
 XX  
 SQ Sequence 79 AA;  
 Query Match 84.5%; Score 93; DB 6; Length 79;  
 Best Local Similarity 94.4%; Pred. No. 1.9e-06;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LFCNVNVCNCFASRNDYS 19  
 DB 24 LFCNVNVCNCFASRNDYS 41  
 RESULT 6  
 AAU75608  
 ID AAU75608 standard; protein; 88 AA.  
 AC  
 AC AAU75608;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.  
 XX  
 XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 82  
 FT /note= "Wild type Cys substituted with Ala"  
 FT  
 XX WO200151523-A2.  
 XX  
 XX 19-JUL-2001.  
 XX  
 XX 08-JAN-2001; 2001WO-US0000565.  
 XX  
 XX 07-JAN-2000; 2000US-00479118.  
 PR  
 XX 04-APR-2000; 2000US-00543371.  
 PR

PR 21-JUL-2000; 2000US-00625191.  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA Kalluri R;  
 XX  
 PI  
 XX WPI; 2002-188037/24.  
 DR  
 XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.  
 PT  
 XX Claim 41; Page 153; 205pp; English.  
 PS  
 XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which  
 CC consists of residues 5-126 of Tumstatin  
 XX  
 SQ Sequence 88 AA;  
 Query Match 84.5%; Score 93; DB 5; Length 88;  
 Best Local Similarity 94.4%; Pred. No. 2.1e-06;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 LFCNVNVCNCFASRNDYS 19  
 DB 34 LFCNVNVCNCFASRNDYS 51  
 RESULT 7  
 AAU75607  
 ID AAU75607 standard; protein; 88 AA.  
 XX  
 AC AAU75607;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 XX Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.  
 DE  
 XX



Db 33 LFCNVNVCNFCASRNDYS 50

||||| ||||| ||||| ||||| |||||

33 LFCNVNVCNFCASRNDYS 50

Best Local Similarity 94.4%; Pred. No. 2.1e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
||||| ||||| ||||| ||||| |||||

Db 33 LFCNVNVCNFCASRNDYS 50

RESULT 10  
AAU75594  
ID AAU75594 standard; protein; 124 AA.  
XX AC AAU75594;  
XX 08-MAY-2002 (first entry)  
XX Human type IV collagen alpha 3 chain mutant, Tumstatin 333.  
DE Human; type IV collagen alpha 3 chain; cytotatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX Homo sapiens.  
OS  
OS  
XX WO200151523-A2.  
XX 19-JUL-2001.  
XX 08-JAN-2001; 2001WO-US000565.  
XX 07-JAN-2000; 2000US-00479118.  
XX 04-APR-2000; 2000US-00543371.  
XX 21-JUL-2000; 2000US-00625191.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.  
XX Example 33; Page; 205pp; English.  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; or (b) by  
XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX endothelial cell apoptosis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; (2) use of  
XX an antibody or peptide that specifically binds the alpha1, alpha2,  
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX preparation of a medicament for inhibiting angiogenesis or cell  
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX of a medicament for treating a proliferative disease in a vertebrate,  
XX where the disease is characterised by angiogenesis that is mediated by  
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX the presence of a medicament for promoting angiogenesis in a tissue; and  
XX (5) use of integrins in the preparation of a medicament for promoting or  
XX inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

33 LFCNVNVCNFCASRNDYS 50

ADA20272 standard; protein; 88 AA.

ADA20272;

20-NOV-2003 (first entry)

Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.

anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
metastasis; basement membrane organisation; type IV collagen network;  
C-terminal globular non-collagenous domain; NC1; type IV collagen;  
cell surface receptor; integrin; angiogenic activity; protein synthesis;  
cytotatic; gene therapy; alpha 3 chain; tumstatin; human;  
tumstatin 5-125-C-A; mutant; mutein.

Synthetic.  
Homo sapiens.

Key Location/Qualifiers  
Misc-difference 81 /notes "Wild-type Cys substituted by Ala at position 125  
of full-length tumstatin"

WO2003059257-A2.

24-JUL-2003.

20-DEC-2002; 2002WO-US040938.

21-DEC-2001; 2001US-00032221.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

Kalluri R;

WPI; 2003-587256/55.

New peptide, useful for preparing a composition for inhibiting tumor  
growth, angiogenic activity or protein synthesis in a mammalian tissue.

Claim 94; SEQ ID NO 34; 240pp; English.

This invention relates to novel isolated proteins and their fragments  
with anti-angiogenic properties. The invention also relates to the DNA  
sequences which encode the novel proteins. A wide variety of diseases are  
the result of undesirable angiogenesis. The formation of new capillaries  
from pre-existing vessels is essential for tumour growth and metastasis.  
Basement membrane organisation is dependent on the assembly of a type IV  
collagen network which may occur through the C-terminal globular non-  
collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
forms are ubiquitously exhibited in human basement membranes. In the  
present invention, cell surface receptors (in particular integrins) which  
specifically bind anti-angiogenic proteins and peptides (in particular  
the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
collagen) are disclosed. The proteins of the invention may inhibit tumour  
growth, angiogenic activity in mammalian tissue or protein synthesis in  
endothelial cells and thus may exhibit cytostatic activity. The DNA  
sequences of the invention may be useful in gene therapy. The present  
sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of  
the "tumstatin" protein of the invention which was derived from the amino  
acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
sequence (Seq ID33) does not appear in the specification but was created  
by the indexer from information given in the specification.

Sequence 88 AA;

Query Match 84.5%; Score 93; DB 6; Length 88;

CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of  
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in  
CC the specification but is derived from the wild type human Tumstatin  
CC sequence given in figure 18A (see AAU75589)

XX  
SQ Sequence 124 AA;

Query Match 84.5%; Score 93; DB 5; Length 124;  
Best Local Similarity 94.4%; Pred. No. 2.9e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
DB 77 LFCNVNVCNCFASRNDYS 94

RESULT 11  
ADA20258  
ID ADA20258 standard; protein; 124 AA.

XX AC ADA20258;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytotatic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.  
XX OS Homo sapiens.

XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 20; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
CC ID20) does not appear in the specification but was created by the indexer  
XX from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 84.5%; Score 93; DB 6; Length 124;  
Best Local Similarity 94.4%; Pred. No. 2.9e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 LFCNVNVCNCFASRNDYS 19  
||||| |||||||  
DB 77 LFCNVNVCNCFASRNDYS 94

RESULT 12  
AAU75597  
ID AAU75597 standard; protein; 132 AA.

XX AC AAU75597;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tum-2.  
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX KW Tumstatin; angiogenesis; tumour; muten; mutant.

XX OS Homo sapiens.  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PR 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and  
XX treating disorders involving angiogenesis.  
XX PS Claim 31; Page 152; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; or (b) by  
XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX endothelial cell apoptosis is mediated by one or more endothelial cell



CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues  
 CC 1-132 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75589)

XX SQ Sequence 132 AA;

Query Match 84.5%; Score 93; DB 5; Length 132;

Best Local Similarity 94.4%; Pred. No. 3.1e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

||||| |||||||  
 Db 78 LFCNVNVCNCFASRNDYS 95

RESULT 13

ADA20261

ID ADA20261 standard; protein; 132 AA.

XX AC ADA20261;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytosstatic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.

XX OS Homo sapiens.

XX PN WQ2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT

PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX Claim 94; SEQ ID NO 23; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-2, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.

XX SQ Sequence 132 AA;

Query Match 84.5%; Score 93; DB 6; Length 132;

Best Local Similarity 94.4%; Pred. No. 3.1e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNCFASRNDYS 19

||||| |||||||  
 Db 77 LFCNVNVCNCFASRNDYS 94

RESULT 14

AAU75596

ID AAU75596 standard; protein; 191 AA.

XX AC AAU75596;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX OS Homo sapiens.

XX PN WQ200151523-A2.

XX PD 19-JUL-2001.

XX PF 08-JAN-2001; 2001WO-US0000565.

XX PR 07-JAN-2000; 2000US-00479118.

XX PR 04-APR-2000; 2000US-00543371.

XX PR 21-JUL-2000; 2000US-00625191.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX DR WPI; 2002-188037/24.

XX PT A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and  
 PT treating disorders involving angiogenesis.

XX PS Example 32; Page; 205pp; English.

XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause

CC apoptosis of endothelial cells. Also described are the following: (1) use

CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for

CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by

CC promoting or inducing endothelial cell apoptosis in a tissue, where the

CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of

CC an antibody or peptide that specifically binds the alpha1, alpha2,

CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the

CC preparation of a medicament for inhibiting angiogenesis or cell

CC proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in the preparation

CC of a medicament for treating a proliferative disease in a vertebrate,

CC where the disease is characterised by angiogenesis that is mediated by

CC receptors to Arresten, Canstatin or Tumstatin and where the receptors

CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one

CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in

CC the presence of a medicament for promoting angiogenesis in a tissue; and

CC (5) use of integrins in the preparation of a medicament for promoting or

CC inducing angiogenesis or cell proliferation in a tissue. The fragments

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

CC or allelic variants are useful in the preparation of a medicament for

CC treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits; or by promoting or

CC inducing endothelial cell apoptosis in a tissue, where the endothelial

CC cell apoptosis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits. The medicament is useful

CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human

CC type IV collagen alpha 3 chain mutant, Tumstatin NS3, which consists of

CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in

CC the specification but is derived from the wild type human Tumstatin

CC sequence given in figure 18A (see AAU75589)

XX SQ Sequence 191 AA;

Query Match 84.5%; Score 93; DB 5; Length 191;

Best Local Similarity 94.4%; Pred. No. 4.5e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 25 LFCNVNVCNCFASRNDYS 42

RESULT 15

ADA20260

ID ADA20260 standard; protein; 191 AA.

XX AC ADA20260;

XX DT 20-NOV-2003 (first entry)

XX DE Human tumstatin deletion protein tum-1 amino acid sequence.

XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

KW metastasis; basement membrane organisation; type IV collagen network;

KW C-terminal globular non-collagenous domain; NC1; type IV collagen;

KW cell surface receptor; integrin; angiogenic activity; protein synthesis;

KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;

XX KW tumstatin NS3.

XX OS Homo sapiens.

XX PN WO2003059257-A2.

XX PD 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Kalluri R;

XX WPI; 2003-587256/55.

XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX PS Claim 94; SEQ ID NO 22; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are

CC the result of undesirable angiogenesis. The formation of new capillaries

CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-

CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the

CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular

CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

CC collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present

CC sequence is that of tum-1 (tumstatin NS3), an abridged form of the

CC "tumstatin" protein of the invention which was derived from the amino

CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This

CC sequence (Seq ID22) does not appear in the specification but was created

CC by the indexer from information given in the specification.

XX SQ Sequence 191 AA;

Query Match 84.5%; Score 93; DB 6; Length 191;

Best Local Similarity 94.4%; Pred. No. 4.5e-06;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19

Db 24 LFCNVNVCNCFASRNDYS 41

Search completed: April 5, 2004, 06:58:32

Job time : 17.1138 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 11.9153 Seconds  
(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNFASTRNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	19	14	US-10-032-221B-41
2	93	84.5	25	14	Sequence 41, Appl
3	93	84.5	27	14	Sequence 37, Appl
4	93	84.5	79	14	Sequence 39, Appl
5	93	84.5	88	14	Sequence 26, Appl
6	93	84.5	124	14	Sequence 33, Appl
7	93	84.5	132	14	Sequence 34, Appl
8	93	84.5	191	14	Sequence 20, Appl
9	93	84.5	211	14	Sequence 22, Appl
10	93	84.5	211	14	Sequence 23, Appl
11	93	84.5	211	14	Sequence 22, Appl
12	93	84.5	232	14	Sequence 46, Appl
13	93	84.5	244	14	Sequence 304, Appl
14	92	83.6	229	14	Sequence 10, Appl
15	92	83.6	229	14	Sequence 302, Appl
					Sequence 2, Appl

16	92	83.6	406	9	US-09-925-302-507	Sequence 507, App
17	92	83.6	1669	15	US-10-372-683-8	Sequence 8, Appl
18	90	81.8	25	14	US-10-032-221B-38	Sequence 38, Appl
19	90	81.8	46	9	US-09-864-761-48095	Sequence 48095, A
20	90	81.8	229	14	US-10-206-699-306	Sequence 306, App
21	90	81.8	309	9	US-09-925-297-496	Sequence 496, App
22	82	74.5	18	14	US-10-206-699-260	Sequence 260, App
23	82	74.5	22	14	US-10-206-699-266	Sequence 266, App
24	81	73.6	18	14	US-10-206-699-259	Sequence 259, App
25	81	73.6	22	14	US-10-206-699-285	Sequence 285, App
26	79	71.8	18	14	US-10-206-699-261	Sequence 261, App
27	79	71.8	22	14	US-10-206-699-267	Sequence 267, App
28	75	68.2	1744	15	US-10-369-493-5832	Sequence 5832, Ap
29	73	66.4	27	14	US-10-032-221B-40	Sequence 40, Appl
30	71	64.5	1759	15	US-10-369-493-7032	Sequence 7032, Ap
31	69	62.7	27	14	US-10-032-221B-42	Sequence 42, Appl
32	64	58.2	142	9	US-09-864-761-38021	Sequence 38021, A
33	64	58.2	228	14	US-10-206-699-307	Sequence 307, App
34	62	56.4	18	14	US-10-206-699-264	Sequence 264, App
35	62	56.4	22	14	US-10-206-699-270	Sequence 270, App
36	61	55.5	14	14	US-10-206-699-3	Sequence 3, Appl
37	61	55.5	15	14	US-10-206-699-210	Sequence 210, App
38	61	55.5	15	14	US-10-206-699-212	Sequence 212, App
39	61	55.5	18	14	US-10-206-699-254	Sequence 254, App
40	61	55.5	20	14	US-10-032-221B-30	Sequence 30, Appl
41	60	54.5	14	14	US-10-206-699-2	Sequence 2, Appl
42	60	54.5	18	14	US-10-206-699-253	Sequence 253, App
43	59	53.6	231	14	US-10-206-699-305	Sequence 305, App
44	58	52.7	14	14	US-10-206-699-4	Sequence 4, Appl
45	58	52.7	18	14	US-10-206-699-255	Sequence 255, App

#### ALIGNMENTS

#### RESULT 1

US-10-032-221B-41  
; Sequence 41, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
; FILE REFERENCE: 2312/20828 (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: TP3 (amino acids 76-94 of SEQ ID NO:10; lysine has been substit  
; OTHER INFORMATION: ed for the phenylalanine residue at position 76 of the full-len  
; OTHER INFORMATION: h-Tumstatin molecule, and cysteine has been substituted for th  
; OTHER INFORMATION: aspartic acid at position 83)  
US-10-032-221B-41

Query Match

100.0%; Score 110; DB 14; Length 19;



Query Match 84.5%; Score 93; DB 14; Length 88;

```

1 CURRENT FILING DATE: 2001-12-21
2 PRIOR APPLICATION NUMBER: FCTY0201/000565
3
4 PRIOR FILING DATE: 2001-01-08
5 PRIOR APPLICATION NUMBER: US 09/625,191
6
7 PRIOR FILING DATE: 2000-07-21
8 PRIOR APPLICATION NUMBER: US 09/543,371
9
10 PRIOR FILING DATE: 2000-04-04
11 PRIOR APPLICATION NUMBER: US 09/479,118
12
13 PRIOR FILING DATE: 2000-01-07
14 PRIOR APPLICATION NUMBER: US 09/335,224
15
16 PRIOR FILING DATE: 1999-06-17
17 PRIOR APPLICATION NUMBER: US 60/126,175
18
19 PRIOR FILING DATE: 1999-03-25
20

```

; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)  
US-10-032-221B-23

Query Match 84.5%; Score 93; DB 14; Length 132;  
Best Local Similarity 94.4%; Pred. No. 8.7e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRDYIS 19  
Db 77 LFCNVNVCNCFASRDYIS 94

## RESULT 9

US-10-032-221B-22  
; Sequence 22, Application US/10032221B  
; Publication No. US2003014481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)  
US-10-032-221B-22

Query Match 84.5%; Score 93; DB 14; Length 191;  
Best Local Similarity 94.4%; Pred. No. 1.2e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRDYIS 19  
Db 24 LFCNVNVCNCFASRDYIS 41

## RESULT 10

US-10-270-877-46  
; Sequence 46, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877

; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-877-46

Query Match 84.5%; Score 93; DB 14; Length 211;  
Best Local Similarity 94.4%; Pred. No. 1.3e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRDYIS 19  
Db 77 LFCNVNVCNCFASRDYIS 94

## RESULT 11

US-10-270-837-46  
; Sequence 46, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-837-46

Query Match 84.5%; Score 93; DB 14; Length 211;  
Best Local Similarity 94.4%; Pred. No. 1.3e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRDYIS 19  
Db 77 LFCNVNVCNCFASRDYIS 94

## RESULT 12

US-10-206-699-304  
; Sequence 304, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 304  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 3 chain  
US-10-206-699-304

Query Match 84.5%; Score 93; DB 14; Length 232;  
Best Local Similarity 94.4%; Pred. No. 1.5e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
||||| |||||||  
Db 65 LFCNVNVCNFCASRNDYS 82  
||||| |||||||

RESULT 13  
US-10-032-221B-10  
; Sequence 10, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram

; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-10

Query Match 84.5%; Score 93; DB 14; Length 244;  
Best Local Similarity 94.4%; Pred. No. 1.5e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
||||| |||||||  
Db 77 LFCNVNVCNFCASRNDYS 94  
||||| |||||||

RESULT 14  
US-10-206-699-302  
; Sequence 302, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Suddaramoorthy, M.  
; APPLICANT: Hudson, B.

; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017

; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 302  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 1 chain  
US-10-206-699-302

Query Match 83.6%; Score 92; DB 14; Length 229;  
Best Local Similarity 88.9%; Pred. No. 2e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
||||| |||||||  
Db 63 LFCNVNVCNFCASRNDYS 80  
||||| |||||||

RESULT 15

US-10-032-221B-2  
; Sequence 2, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram

; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-2

Query Match 83.6%; Score 92; DB 14; Length 229;  
Best Local Similarity 88.9%; Pred. No. 2e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNFCASRNDYS 19  
||||| |||||||  
Db 63 LFCNVNVCNFCASRNDYS 80  
||||| |||||||

Search completed: April 5, 2004, 07:36:07  
Job time : 11.9153 secs



us-10-032-221b-41.rapb

Mon Apr 5 07:53:15 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 4.46247 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221B-41

Perfect score: 110

Sequence: 1 KLFCNVNVCNPFASNDYS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	84.5	211	4	US-09-512-563C-46
2	93	84.5	218	2	US-08-399-889-25
3	93	84.5	218	3	US-09-167-364-25
4	93	84.5	218	3	US-09-439-897-4
5	93	84.5	268	4	US-09-589-927-6
6	93	84.5	268	4	US-09-277-665-6
7	93	84.5	268	4	US-09-589-987-6
8	92	83.6	260	4	US-09-589-927-2
9	92	83.6	260	4	US-09-277-665-2
10	92	83.6	260	4	US-09-589-987-2
11	92	83.6	471	2	US-08-399-889-24
12	92	83.6	471	3	US-09-167-364-24
13	92	83.6	471	3	US-09-439-897-2
14	90	81.8	264	4	US-09-589-927-10
15	90	81.8	264	4	US-09-277-665-10
16	90	81.8	264	4	US-09-589-987-10
17	64	58.2	260	4	US-09-589-927-12
18	64	58.2	260	4	US-09-277-665-12
19	64	58.2	260	4	US-09-589-987-12
20	62	56.4	1694	1	US-08-494-168-2
21	59	53.6	260	4	US-09-589-927-8
22	59	53.6	260	4	US-09-277-665-8
23	59	53.6	260	4	US-09-589-987-8
24	58	52.7	258	4	US-09-589-927-4
25	58	52.7	258	4	US-09-277-665-4
26	58	52.7	258	4	US-09-589-987-4
27	49	44.5	347	4	US-09-636-215-590

28 49 44.5 347 4 US-09-685-166A-590 Sequence 590, Appl  
29 49 44.5 374 2 US-08-820-170A-25 Sequence 25, Appl  
30 49 44.5 374 3 US-09-055-899-25 Sequence 25, Appl  
31 49 44.5 374 3 US-09-273-565-25 Sequence 25, Appl  
32 49 44.5 374 4 US-09-565-538-25 Sequence 25, Appl  
33 49 44.5 374 4 US-09-661-468-25 Sequence 25, Appl  
34 49 44.5 374 4 US-09-976-165-25 Sequence 25, Appl  
35 49 44.5 374 4 US-09-227-853A-2 Sequence 2, Appl  
36 49 44.5 374 5 PCT-US95-06385-2 Sequence 2, Appl  
37 45 40.0 69 4 US-09-621-976-5669 Sequence 5669, Ap  
38 44 40.0 49 1 US-07-865-166A-6 Sequence 6, Appl  
39 42 38.2 45 3 US-08-965-903B-19 Sequence 19, Appl  
40 42 38.2 100 3 US-08-965-903B-11 Sequence 11, Appl  
41 42 38.2 100 4 US-09-370-398-7 Sequence 7, Appl  
42 42 38.2 100 4 US-10-030-190-7 Sequence 7, Appl  
43 42 38.2 124 3 US-08-965-903B-4 Sequence 4, Appl  
44 42 38.2 182 4 US-09-543-681A-6443 Sequence 6443, Ap  
45 42 38.2 461 1 US-08-385-229-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-512-563C-46  
; Sequence 46, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512.563C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-09-512-563C-46

Query Match 84.5%; Score 93; DB 4; Length 211;  
Best Local Similarity 94.4%; Pred. No. 1.9e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 LFCNVNVCNPFASNDYS 19  
DB 77 LFCNVNVCNPFASNDYS 94

RESULT 2  
US-08-399-889-25  
; Sequence 25, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399.889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patentin ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 218  
; TYPE: PRT

; ORGANISM: Human  
US-08-399-889-25

Query Match 84.5%; Score 93; DB 2; Length 218;  
Best Local Similarity 94.4%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19  
Db 51 LFCNVNVCNCFASRDY 68

## RESULT 3

US-09-167-364-25  
; Sequence 25, Application US/09167364  
; Patent No. 6007980

## ; GENERAL INFORMATION:

; APPLICANT: Readers, Stephen T

; APPLICANT: Morrison, Karen E

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 951263B

; CURRENT APPLICATION NUMBER: US/09/167,364

; CURRENT FILING DATE: 1998-10-07

; EARLIER APPLICATION NUMBER: 08/399889

; EARLIER FILING DATE: 1995-03-07

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 25

; LENGTH: 218

; TYPE: PRT

; ORGANISM: Human

US-09-167-364-25

Query Match 84.5%; Score 93; DB 3; Length 218;  
Best Local Similarity 94.4%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19  
Db 51 LFCNVNVCNCFASRDY 68

## RESULT 4

US-09-439-897-4

; Sequence 4, Application US/09439897

; Patent No. 6277558

## ; GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 95-1263-C

; CURRENT APPLICATION NUMBER: US/09/439,897

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 218

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-439-897-4

Query Match 84.5%; Score 93; DB 3; Length 218;  
Best Local Similarity 94.4%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19  
Db 51 LFCNVNVCNCFASRDY 68

## RESULT 5

US-09-589-927-6

; Sequence 6, Application US/09589927

; Patent No. 6432706  
; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center

; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions

; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,927

; CURRENT FILING DATE: 2000-06-07

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 268

; TYPE: PRT

; ORGANISM: Human

US-09-589-927-6

Query Match 84.5%; Score 93; DB 4; Length 268;  
Best Local Similarity 94.4%; Pred. No. 2.5e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19  
Db 101 LFCNVNVCNCFASRDY 118

## RESULT 6

US-09-277-665-6

; Sequence 6, Application US/09277665

; Patent No. 6440729

## ; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center

; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions

; FILE REFERENCE: 94525-1

; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 268

; TYPE: PRT

; ORGANISM: Human

US-09-277-665-6

Query Match 84.5%; Score 93; DB 4; Length 268;  
Best Local Similarity 94.4%; Pred. No. 2.5e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LFCNVNVCNCFASRDY 19  
Db 101 LFCNVNVCNCFASRDY 118

## RESULT 7

US-09-589-987-6

; Sequence 6, Application US/09589987

; Patent No. 6498140

## ; GENERAL INFORMATION:

; APPLICANT: University of Kansas Medical Center

; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions

; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,987

; CURRENT FILING DATE: 2000-06-07

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 268

; TYPE: PRT

; ORGANISM: Human

US-09-589-987-6

Query Match 84.5%; Score 93; DB 4; Length 268;

Best Local Similarity 94.4%; Pred. No. 2.5e-06; Mismatches 1; Indels 0; Gaps 0;  
Matches 17; Conservative 0;  
QY 2 LFCNVNVCNFAASNDYS 19  
Db 101 LFCNVNVCNFAASNDYS 118  
||||| |||||||

RESULT 8  
US-09-589-927-2  
; Sequence 2, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-2

Query Match 83.6%; Score 92; DB 4; Length 260;  
Best Local Similarity 88.9%; Pred. No. 3.3e-06; Mismatches 1; Indels 0; Gaps 0;  
Matches 16; Conservative 1;

QY 2 LFCNVNVCNFAASNDYS 19  
Db 94 LFCNINNVCNFAASNDYS 111  
||||| |||||||

RESULT 9  
US-09-277-665-2  
; Sequence 2, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251-I  
; CURRENT APPLICATION NUMBER: US/09/277,665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-2

Query Match 83.6%; Score 92; DB 4; Length 260;  
Best Local Similarity 88.9%; Pred. No. 3.3e-06; Mismatches 1; Indels 0; Gaps 0;  
Matches 16; Conservative 1;

QY 2 LFCNVNVCNFAASNDYS 19  
Db 94 LFCNINNVCNFAASNDYS 111  
||||| |||||||

RESULT 10  
US-09-589-987-2  
; Sequence 2, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251

; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-2

Query Match 83.6%; Score 92; DB 4; Length 260;  
Best Local Similarity 88.9%; Pred. No. 3.3e-06; Mismatches 1; Indels 0; Gaps 0;  
Matches 16; Conservative 1;

QY 2 LFCNVNVCNFAASNDYS 19  
Db 94 LFCNINNVCNFAASNDYS 111  
||||| |||||||

RESULT 11  
US-08-399-889-24  
; Sequence 24, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Calf  
US-08-399-889-24

Query Match 83.6%; Score 92; DB 2; Length 471;  
Best Local Similarity 88.9%; Pred. No. 6.1e-06; Mismatches 1; Indels 0; Gaps 0;  
Matches 16; Conservative 1;

QY 2 LFCNVNVCNFAASNDYS 19  
Db 304 LFCNINDVCNFAASNDYS 321  
||||| |||||||

RESULT 12  
US-09-167-364-24  
; Sequence 24, Application US/09167364  
; Patent No. 6007980  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263B  
; CURRENT APPLICATION NUMBER: US/09/167,364  
; CURRENT FILING DATE: 1998-10-07  
; EARLIER APPLICATION NUMBER: 08/399889  
; EARLIER FILING DATE: 1995-03-07  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Calf  
US-09-167-364-24

Query Match 83.6%; Score 92; DB 3; Length 471;

Best Local Similarity 88.9%; Pred. No. 6.1e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 LFCNVNVCNCFASRNDYS 19  
Db 304 LFCNINDVCNCFASRNDYS 321

## RESULT 13

US-09-439-897-2  
; Sequence 2, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-09-439-897-2

Query Match 83.6%; Score 92; DB 3; Length 471;  
Best Local Similarity 88.9%; Pred. No. 6.1e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 LFCNVNVCNCFASRNDYS 19  
Db 304 LFCNINDVCNCFASRNDYS 321

## RESULT 14

US-09-589-927-10  
; Sequence 10, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525I  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-10

Query Match 81.8%; Score 90; DB 4; Length 264;  
Best Local Similarity 83.3%; Pred. No. 6.7e-06;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 LFCNVNVCNCFASRNDYS 19  
Db 98 MFCNINNVCNCFASRNDYS 115

## RESULT 15

US-09-277-665-10  
; Sequence 10, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-I  
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-10

Query Match 81.8%; Score 90; DB 4; Length 264;  
Best Local Similarity 83.3%; Pred. No. 6.7e-06;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFCNVNVCNCFASRNDYS 19  
Db 98 MFCNINNVCNCFASRNDYS 115

Search completed: April 5, 2004, 07:07:26  
Job time : 5.45247 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:51:17 ; Search time 5.5569 Seconds

(without alignments)  
467.378 Million cell updates/sec

Title: US-10-032-221b-42  
Perfect score: 145  
Sequence: 1 QRFTTMEFLFDVNDVNFASRNDYS 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:  
1: PIR1:  
2: PIR2:  
3: PIR3:  
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	84.1	220	2 B49736	collagen alpha 3(I
2	122	84.1	1670	1 CGHU3B	collagen alpha 3(I
3	121	83.4	471	2 A39024	collagen alpha 3(I
4	116	80.0	161	2 S49488	collagen alpha 3(I
5	116	80.0	246	2 I48302	collagen alpha 3(I
6	106	73.1	253	2 I48304	collagen alpha 5(I
7	106	73.1	754	2 A55267	collagen alpha 5(I
8	106	73.1	1691	1 S22917	collagen alpha 5(I
9	105	72.4	258	2 B61228	collagen alpha 1(I
10	105	72.4	1669	1 CGHU4B	collagen alpha 1(I
11	105	72.4	1669	1 CGM54B	collagen alpha 1(I
12	96	66.2	1747	2 A54121	collagen alpha 1(I
13	96	66.2	1752	2 A45407	collagen alpha 1(I
14	91	62.8	1758	2 T29350	hypothetical prote
15	91	62.8	1759	2 T29351	collagen alpha 2(I
16	91	62.8	1753	2 S16366	collagen alpha 2(I
17	87	60.0	281	2 A44476	collagen alpha 2(I
18	83	57.2	1744	2 S40991	collagen alpha 1(I
19	81	55.9	1691	1 CGHU6B	collagen alpha 6(I
20	78	53.8	775	2 A61228	collagen alpha 2(I
21	78	53.8	1707	2 A33526	collagen alpha 2(I
22	78	53.8	1712	1 CGHU2B	collagen alpha 2(I
23	65	44.8	1761	2 T13990	collagen type IV a
24	64	44.1	312	2 I48303	collagen alpha 4(I
25	64	44.1	623	2 A45137	collagen alpha 4(I
26	64	44.1	1690	1 CGHU1B	collagen alpha 4(I
27	64	44.1	1775	2 A31893	collagen alpha 1(I
28	63	43.4	453	2 S18804	collagen alpha 4(I
29	56	38.6	332	2 P82140	C4-dicarbonylate-b

30 55 37.9 331 2 H83000 probable C4-dicarb  
31 52 35.9 457 2 T23494 phenylalanine 4-mo  
32 51.5 35.5 610 2 C70126 DNA mismatch repai  
33 50 34.5 331 2 A83534 probable C4-dicarb  
34 47.5 32.8 334 2 C71718 hypothetical prote  
35 47.5 32.8 490 2 T24497 hypothetical prote  
36 47 32.4 252 2 S50806 hypothetical prote  
37 46.5 32.1 334 2 B97715 hypothetical prote  
38 46 31.7 310 2 T01266 starch synthase DU  
39 46 31.7 351 2 JCS904 major capsid prote  
40 46 31.7 364 2 A97335 probable membrane  
41 46 31.7 795 2 AF2444 hypothetical prote  
42 46 31.7 1124 2 F71719 hypothetical prote  
43 46 31.7 1462 2 T42639 glucocorticoid rec  
44 46 31.7 1463 2 T30193 nuclear receptor c  
45 46 31.7 1674 2 T01265 starch synthase DU

## ALIGNMENTS

## RESULT 1

B49736  
collagen alpha 3(IV) chain, medium splice form - human (fragment)  
N:Contains: collagen alpha 3(IV) chain, splice form GP-V  
C:Species: Homo sapiens (man)  
C>Date: 03-May-1994 #sequence revision 12-Nov-1999 #text\_change 17-Mar-2000  
C:Accession: B49736; D49736; S69111  
R:Feng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A:Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.  
A:Reference number: A49736; MUID:94124597; PMID:8294492  
A:Accession: B49736  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 169-220 <FEN1>  
A:Accession: D49736  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 22-220 <FEN2>  
A:Cross-references: GB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107  
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank  
R:Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; W.  
Eur. J. Biochem. 229, 754-760, 1995  
A:Title: Characterization and expression of multiple alternatively spliced transcripts  
A:Note: This is the conceptual translation of the nucleic acid submitted to GenBank  
A:Reference number: S69111; MUID:95278230; PMID:7758473  
A:Accession: S69111  
A:Molecule type: mRNA  
A:Residues: 1-45,169-204,'L',206-220 <PEN>  
C:Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.  
C:Genetics:  
A:Gene: GDB:COL4A3  
A:Cross-references: GDB:128351; OMIM:120070  
A:Map position: 2q36-2q37  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrac  
F:1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pre  
F:1-45,169-220/Product: collagen alpha 3(IV) chain, splice from GP-V (fragment) #status  
F:22-220/Domain: carboxyl-terminal nonhelical, NCI <NC1>  
F:34-134/Domain: collagen IV carboxyl-terminal repeat <CTL>

Query Match 84.1% Score 122; DB 2; Length 220;  
Best Local Similarity 92.3%; Pred. No. 7.4e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTTMEFLFDVNDVNFASRNDYS 27

DB 78 QRFTTMEFLFDVNDVNFASRNDYS 103

RESULT 2  
CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human  
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
 C/Species: Homo sapiens (man)  
 C/Date: 28-Oct-1994 #sequence\_revision 03-Oct-1995 #text change 22-Jun-1999  
 C/Accession: A54763; A43928; A44043; A45971; A39786  
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reuters, S.T.  
 J. Biol. Chem. 269, 23013-23017, 1994  
 A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
 A/Reference number: A54763; MUID:94364994; PMID:8083201  
 A/Accession: A54763  
 A/Molecule type: mRNA  
 A/Residues: 1-1670 <MAR>  
 A/Cross-references: GB:XB0031; NID:G577563; PID:G577564  
 A/Experimental source: kidney  
 R/Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
 J. Clin. Invest. 89, 592-601, 1992  
 A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha  
 A/Reference number: A43928; MUID:92147878; PMID:1737849  
 A/Accession: A43928  
 A/Molecule type: mRNA  
 A/Residues: 1331-1524, 'I', 1526-1670 <TUR>  
 A/Cross-references: GB:M61379  
 A/Experimental source: kidney  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 267, 19780-19784, 1992  
 A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture  
 ction.  
 A/Reference number: A44043; MUID:93015826; PMID:1400291  
 A/Accession: A44043  
 A/Molecule type: DNA, mRNA  
 A/Residues: 1386-1670 <QUI>  
 A/Cross-references: GB:N22993; NID:G177895; PIDN:AAA1610.1; PID:G177896  
 A/Note: sequence extracted from NCBI backbone (NCBI:1115597)  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 269, 17358, 1994  
 A/Reference number: A44738; MUID:94274734; PMID:8006044  
 A/Contents: annotation; erratum; correction to intronic sequence in A44043  
 R/Bernal, D.; Quinones, S.; Saus, J.  
 J. Biol. Chem. 268, 12090-12094, 1993  
 A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
 A/Reference number: A45971; MUID:93280184; PMID:8505332  
 A/Accession: A45971  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1427-1444 <BER>  
 A/Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly ident  
 R/Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reuters, S.T.  
 Am. J. Hum. Genet. 49, 545-554, 1991  
 A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
 A/Reference number: A39786; MUID:91353570; PMID:1882840  
 A/Accession: A39786  
 A/Molecule type: mRNA  
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>  
 A/Cross-references: GB:S5790; NID:G234418; PIDN:AA19637.1; PID:G234419  
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
 C/Genetics:  
 A/Gene: GDB:COL4A3  
 A/Cross-references: GDB:128351; OMIM:120070  
 A/Map position: 2q36-2q37  
 A/Introns: 1385/1, 1418/1, 1486/1, 1547/2, 1585/3, 1643/2 #status incomplete  
 A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l  
 C/Function:  
 A/Description: minor structural component of extracellular basement membrane in kidney g  
 C/Superfamily: collagen alpha 1(IV) chain  
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:23-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>  
 F:23-42/Domain: amino-terminal nonhelical, NH1 <NHI>

F:43-1438/Region: interrupted helical  
 F:791-793/Region: cell attachment (R-G-D) motif  
 F:996-998/Region: cell attachment (R-G-D) motif  
 F:1154-1156/Region: cell attachment (R-G-D) motif  
 F:1306-1308/Region: cell attachment (R-G-D) motif  
 F:1345-1347/Region: cell attachment (R-G-D) motif  
 F:1432-1434/Region: cell attachment (R-G-D) motif  
 F:1439-1470/Domain: carboxyl-terminal nonhelical, NC1 <NCL>  
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
 F:31.33.39.41.125.422.476.479.682.722.809.1387/Disulfide bonds: interchain #status pred  
 F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:1450-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F:1505-1511, 1616-1622/Disulfide bonds: #status predicted  
 F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
 Query Match 84.1%; Score 122; DB 1; Length 1670;  
 Best Local Similarity 92.3%; Pred. No. 7.7e-10;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27  
 DB 1495 QRFTHMPFLFDNVNDVNFASRNDYS 1520  
 RESULT 3  
 A39024  
 collagen alpha 3(IV) chain - bovine (fragment)  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text change 13-Aug-1999  
 C/Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815  
 R/Morrison, K.E.; Germino, G.G.; Reuters, S.T.  
 J. Biol. Chem. 266, 34-39, 1991  
 A/Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
 A/Reference number: A39024; MUID:91093146; PMID:1985905  
 A/Accession: A39024  
 A/Molecule type: mRNA  
 A/Residues: 1-471 <MOR>  
 A/Cross-references: EMBL:MG3139; NID:G162886; PIDN:AAA62708.1; PID:G162887  
 R/Butkowski, R.J.; Wieslander, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
 J. Biol. Chem. 262, 7874-7877, 1987  
 A/Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
 A/Reference number: S18432; MUID:87222419; PMID:2438283  
 A/Accession: S20672  
 A/Molecule type: protein  
 A/Residues: 227-228, 'X', 230-244 <BUT>  
 R/Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
 J. Biol. Chem. 263, 13374-13380, 1988  
 A/Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
 A/Reference number: S17802; MUID:88330844; PMID:3417661  
 A/Accession: S17802  
 A/Molecule type: protein  
 A/Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
 R/Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
 J. Biol. Chem. 265, 5466-5469, 1990  
 A/Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
 A/Reference number: A35167; MUID:90202779; PMID:2318822  
 A/Accession: A35167  
 A/Molecule type: protein  
 A/Residues: 236-258 <GUN>  
 R/Gunwar, S.; Ballerster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; NC  
 J. Biol. Chem. 266, 15318-15324, 1991  
 A/Title: Glomerular basement membrane. Identification of dimeric subunits of the noncol  
 A/Reference number: A39419; MUID:91332055; PMID:1869555  
 A/Accession: C39419  
 A/Molecule type: protein  
 A/Residues: 236-255 <GU2>  
 C/Superfamily: collagen alpha 1(IV) chain  
 C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;  
 F:1-238/Domain: collagenous (fragment) #status predicted <COL>  
 F:233-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NCL>  
 F:233-353/Domain: repeat NC1 #status predicted <NC11>  
 F:354-471/Domain: repeat NC1 #status predicted <NC12>





A; Cross-references: GB:S75903; NID:G913892; PIDN:AAB33374.1.; PID:G913893  
A; Note: premature termination mutation from a patient with Alport syndrome; one other  
R; Lemlink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.  
Genomics 17, 485-489, 1993

A; Title: Identification of four novel mutations in the COL4A5 gene of patients with Al;  
A; Reference number: I54188; MUID:94010948; PMID:8406498  
A; Accession: I54188  
A; Status: translated from GB/ENBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 1604-1607, "VHDAYKC" <LEM>  
A; Cross-references: GB:S65767; NID:G425563; PIDN:AAD13967.1.; PID:G4261667  
A; Note: frameshift mutation from a patient with Alport syndrome; five other mutations;  
C; Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.

C; Genetics:  
A; Gene: GDB:COL4A5; ATS  
A; Cross-references: GDB:120596; OMIM:303630  
A; Map position: Xq22-Xq22  
A; Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 22;  
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1165/1;  
A; Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands wi  
C; Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
mer trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-  
C; Function:  
C; Description: minor structural component of extracellular basement membrane  
C; Superfamily: collagen alpha 1(IV) chain  
C; Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; gly  
F; 1-26/Domain: signal sequence #status predicted <Sig>  
F; 27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MA  
F; 27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #statu  
F; 27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>  
F; 42-1462/Region: interrupted helical  
F; 1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F; 1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F; 1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F; 23,32,38,40,120,451,481,484/Disulfide bonds: interchain #status predicted  
F; 125/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F; 1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicte  
F; 1527-1533,1638-1644/Disulfide bonds: #status predicted  
F; 1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicte

Query Match 73.1%; Score 106; DB 1; Length 1691;  
Best Local Similarity 73.1%; Pred. No. 2e-07;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTWPFLFDNVNDNFASRNDYS 27  
:|||||:|:|:|||||

Dd 1517 RRSTWPFMFCNNVCNPFASRNDYS 1542  
:|||||:|:|:|||||

RESULT 9  
B61228  
Collagen alpha 1(IV) chain - rabbit (fragment)  
C; Species: Oryctolagus cuniculus (domestic rabbit)  
C; Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 17-Mar-1999  
C; Accession: B61228  
R; Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.  
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991  
A; Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothel  
A; Reference number: A61228; MUID:92010685; PMID:1717398  
A; Accession: B61228  
A; Status: preliminary  
A; Molecule type: mRNA  
A; Residues: 1-258 <YAM>  
C; Superfamily: collagen alpha 1(IV) chain

Query Match 72.4%; Score 105; DB 2; Length 258;  
Best Local Similarity 73.1%; Pred. No. 3.2e-08;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRTTWPFLFDNVNDNFASRNDYS 27  
:|||||:|:~|:|||||

Db 84 RKFTWPELFNCINNVCFASRNDYS 109

## RESULT 10

CGHU4B

collagen alpha 1(IV) chain precursor - human  
N;Alternate names: procollagen alpha 1(IV) chain  
C;Species: Homo sapiens (man)  
C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58  
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll  
A;Reference number: S16876; MUID:89340433; PMID:2701944  
A;Accession: S16876  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1669 <SOI1>  
A;Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803  
R;Soininen, R.; Huotari, M.; Hosikaka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A;Reference number: A92650; MUID:89034231; PMID:3182844  
A;Accession: A32117  
A;Molecule type: DNA  
A;Residues: 1-28 <SOI2>  
A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R;Poeschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c  
A;Reference number: S02738; MUID:89030632; PMID:2846280  
A;Accession: S02738  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-6 'L', 8-28 <POE>  
A;Cross-references: EMBL:X12784; NID:G30072  
R;Brazier, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;  
Eur. J. Biochem. 168, 529-536, 1987  
A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem  
A;Reference number: S00048; MUID:88029471; PMID:3311751  
A;Accession: S00048  
A;Molecule type: mRNA  
A;Residues: 1-318, 'A', 320-944 <BRAL>  
A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A;Accession: S25826  
A;Molecule type: protein  
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S  
A;Reference number: A23115; MUID:86004708; PMID:4043082  
A;Accession: A23115  
A;Molecule type: protein  
A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A;Experimental source: Placenta  
A;Note: the amino end of the mature form is blocked  
R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1987  
A;Title: Complete primary structure of the alpha1(1)-chain of human basement membrane (ty  
A;Reference number: S00207; MUID:88083584; PMID:3691802  
A;Accession: S00207  
A;Molecule type: mRNA  
A;Residues: 244-530 <SOI3>  
A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen  
A;Reference number: S39614; MUID:94038963; PMID:8223488  
A;Accession: S39614  
A;Molecule type: protein  
A;Residues: 371-554 <EBL>  
R;Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid  
A;Reference number: A02863; MUID:85003629; PMID:6434307  
A;Accession: A02863  
A;Molecule type: protein  
A;Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 99  
A;Experimental source: Placenta  
R;Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A;Title: Peppin fragments of human placental basement-membrane collagens showing inter  
A;Reference number: S16908; MUID:82005835; PMID:6792033  
A;Accession: A58517  
A;Molecule type: protein  
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-  
R;MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Piezsek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A;Title: Isolation and characterization of pepsin-solubilized human basement membrane  
A;Reference number: S16910; MUID:84053346; PMID:6416291  
A;Accession: S16910  
A;Molecule type: protein  
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'V', 946, 'X', 94  
A;Experimental source: Placenta  
R;Philajantem, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.;  
J. Biol. Chem. 260, 7681-7687, 1985  
A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen  
A;Reference number: S01466; MUID:85207819; PMID:2581969  
A;Accession: S01466  
A;Molecule type: mRNA  
A;Residues: 1256-1669 <PIH>  
A;Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424  
R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A;Title: Restricted homology between human alpha-1 type IV and other procollagen chain  
A;Reference number: S16879; MUID:85216555; PMID:2582422  
A;Accession: S16879  
A;Molecule type: mRNA  
A;Residues: 1259-1669 <BRI>  
A;Cross-references: EMBL:M1315; NID:G180817; PIDN:AAA52042.1; PID:G180818  
R;Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss  
Eur. J. Biochem. 147, 217-224, 1985  
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha  
A;Reference number: A02864; MUID:85127033; PMID:2578961  
A;Accession: S19091  
A;Molecule type: protein  
A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491; 1501-1514, 'X', 1516-1519; 1534-1553, 'X'  
R;Siebold, B.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 176, 617-624, 1988  
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyte:  
A;Reference number: S02550; MUID:89005112; PMID:2844531  
A;Contents: annotation; disulfide bonds  
C;Genetics:  
A;Gene: GDB:COL4A1  
A;Cross-references: GDB:119791; OMIM:120130  
A;Map position: 13q34-13q34  
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 23;  
/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 1;  
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 1;  
ociations among trimer amino-terminal domains (disulfide and desmosine cross-links), d;  
x-trimer associations in the interrupted helical domain (with disulfide and desmosine)  
C;Function:  
A;Description: structural component of extracellular basement membrane  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicati  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>  
F;163-1440/Domain: interrupted helical <COL>  
F;414-452/Region: integrin binding #status experimental  
F;597-599/Region: cell attachment (R-G-D) motif  
F;917-919/Region: cell attachment (R-G-D) motif  
F;968-970/Region: cell attachment (R-G-D) motif  
F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <Ctri>



RESULT 13

A45407  
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C:Accession: A54407; A43903; A33940  
E:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.  
J. Biol. Chem. 268, 5249-5254, 1993  
A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and  
A:Reference number: A5407; MUID:9318684; PMID:8444899

```

Qy      2  QRFTMPFLFNDVNDVNFA SRNDYS  27
      ||||| : |||||
Db      1581 QRFTMPFLFCDFNVNCVYASRNDKS  1606

```

```

Query Match      62.8%; Score 91; DB 2; Length 1759;
Best Local Similarity 69.2%; Pred. No. 3.8e-05;
Matches 18; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy  2  QRRTMPFLFDNVNDVDFNFASRNDYS  27
      ||||| : : : |||||
Db  1582 QRSTMPFLFDNFNVNVCNYSRNDKS  1607

```

Search completed: April 5, 2004, 07:05:39  
Job time : 5.5569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:45:17 ; Search time 3.3952 Seconds  
(without alignments)  
413.557 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTTPFLFDVNDVNFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	122	84.1	1670	1 CA34 HUMAN	Q01955 homo sapien
2	121	83.4	471	1 CA34 BOVIN	Q28084 bos taurus
3	106	73.1	754	1 CA54 CANFA	P29407 canis faml
4	106	73.1	1685	1 CA54 HUMAN	P29400 homo sapien
5	105	72.4	1669	1 CA14 HUMAN	P02462 homo sapien
6	105	72.4	1669	1 CA14 MOUSE	P02463 mus musculus
7	91	62.8	1763	1 CA24 ASCSU	P27393 ascaris suu
8	87	60.0	1758	1 CA24 CAEEL	P17140 caenorhabdi
9	83	57.2	1758	1 CA14 CAEEL	P17139 caenorhabdi
10	81	55.9	1691	1 CA64 HUMAN	Q14031 homo sapien
11	78	53.8	1707	1 CA24 MOUSE	P08122 mus musculus
12	78	53.8	1712	1 CA24 HUMAN	P08572 homo sapien
13	64	44.1	623	1 CA44 RABIT	P55787 cryptolagus
14	64	44.1	1690	1 CA44 HUMAN	P53420 homo sapien
15	64	44.1	1775	1 CA14 DROME	P08120 drosophila
16	63	43.4	453	1 CA44 BOVIN	Q29442 bos taurus
17	52	35.9	371	1 CYB LIAPA	O48093 liasis macu
18	52	35.9	371	1 CYB LIAPA	O48098 liasis papu
19	52	35.9	457	1 PH4H CAEEL	P90925 caenorhabdi
20	51.5	35.5	610	1 MUTL BORBU	O51229 borrelia bu
21	49	33.8	371	1 CYB LOXBI	O48100 loxocemus b
22	48	33.1	379	1 CYB DIPNE	Q98400 dipodomys n
23	47.5	32.6	334	1 Y092 RICPR	Q98555 rickettsia
24	47	32.4	252	1 YUG6 YEAST	P40364 saccharomyc
25	47	32.4	371	1 CYB ERYJA	O48076 eryx jaculu
26	47	32.4	825	1 XFP BIFAN	Q9AEM9 bifidobacte
27	46	31.7	218	1 IF6 METAC	Q8TIN4 methanosarc
28	46	31.7	328	1 NODZ AZOQA	O43966 azorhizobiu
29	46	31.7	371	1 CYB COLCO	Q9M110 coluber con
30	46	31.7	371	1 CYB ELANI	Q9MLX8 elapsoides
31	46	31.7	1462	1 NCO2 MOUSE	Q61026 mus musculu
32	46	31.7	1464	1 NCO2 HUMAN	Q15596 homo sapien
33	46	31.7	1465	1 NCO2 RAT	Q9WUI9 rattus norv

## RESULT 1

CA34 HUMAN  
ID CA34\_HUMAN STANDARD; PRT; 1670 AA.  
AC Q01955; Q9BQ2; 125 1 YOM2 PHOPR  
DT 01-OCT-1996 (Rel. 34, Created) 31.4 125 1 YOM2 PHOPR  
DT 15-JUL-1999 (Rel. 38, Last sequence update) 31.4 288 1 T2D2\_STRPN  
DT 10-OCT-2003 (Rel. 42, Last annotation update) 1397 1 CID\_DROME  
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen). 370 1 CYB\_MICIK  
GN COL4A3. 430 1 RUMA\_SALTI  
OS Homo sapiens (Human). 430 1 RUMA\_SALTY  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; 430 1 RUMA\_ECOL6  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 430 1 RUMA\_ECOLI  
OX NCBI\_TaxID=9606; 430 1 RUMA\_ECOLI  
RN [1] 430 1 RUMA\_ECOLI  
RP SEQUENCE FROM N.A. 430 1 RUMA\_ECOLI  
RC TISSUE=Kidney; 430 1 RUMA\_ECOLI  
RX MEDLINE=94364994; PubMed=8083201; 430 1 RUMA\_ECOLI  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.; 430 1 RUMA\_ECOLI  
RT "Complete primary structure of the human alpha 3(IV) collagen chain. 430 1 RUMA\_ECOLI  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in 430 1 RUMA\_ECOLI  
RL human tissues."; 430 1 RUMA\_ECOLI  
RL J. Biol. Chem. 269:23013-23017(1994). 430 1 RUMA\_ECOLI  
RN [2] 430 1 RUMA\_ECOLI  
RP REVISIONS. 430 1 RUMA\_ECOLI  
RA Leinonen A.; 430 1 RUMA\_ECOLI  
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases. 430 1 RUMA\_ECOLI  
[3] 430 1 RUMA\_ECOLI  
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167; 430 1 RUMA\_ECOLI  
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND 430 1 RUMA\_ECOLI  
RP CYS-1661; AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451; 430 1 RUMA\_ECOLI  
RP PRO-574; GLU-1269 AND PRO-1474. 430 1 RUMA\_ECOLI  
RX MEDLINE=21064696; PubMed=11134255; 430 1 RUMA\_ECOLI  
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G., 430 1 RUMA\_ECOLI  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.; 430 1 RUMA\_ECOLI  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in 430 1 RUMA\_ECOLI  
RT autosomal Alport syndrome."; 430 1 RUMA\_ECOLI  
RL J. Am. Soc. Nephrol. 12:97-106(2001). 430 1 RUMA\_ECOLI  
RN [4] 430 1 RUMA\_ECOLI  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE. 430 1 RUMA\_ECOLI  
RX MEDLINE=93015826; PubMed=1400291; 430 1 RUMA\_ECOLI  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.; 430 1 RUMA\_ECOLI  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the 430 1 RUMA\_ECOLI  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially 430 1 RUMA\_ECOLI  
RT antigenic region at the triple helix/NC1 domain junction."; 430 1 RUMA\_ECOLI  
RL J. Biol. Chem. 267:19780-19784(1992). 430 1 RUMA\_ECOLI  
RN [5] 430 1 RUMA\_ECOLI  
RP SEQUENCE OF 1453-1670 FROM N.A. 430 1 RUMA\_ECOLI  
RX MEDLINE=91353570; PubMed=1882840; 430 1 RUMA\_ECOLI  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.; 430 1 RUMA\_ECOLI  
RT "Sequence and localization of a partial cDNA encoding the human alpha 430 1 RUMA\_ECOLI  
RT 3 chain of type IV collagen."; 430 1 RUMA\_ECOLI  
RL Am. J. Hum. Genet. 49:545-554(1991). 430 1 RUMA\_ECOLI  
RN [6] 430 1 RUMA\_ECOLI  
RP SEQUENCE OF 1331-1670 FROM N.A. 430 1 RUMA\_ECOLI  
RC TISSUE=Kidney; 430 1 RUMA\_ECOLI  
RX MEDLINE=92147878; PubMed=1737849; 430 1 RUMA\_ECOLI

## ALIGNMENTS

34 45.5 31.4 125 1 YOM2 PHOPR  
35 45.5 31.4 288 1 T2D2\_STRPN  
36 45.5 31.4 1397 1 CID\_DROME  
37 45 31.0 370 1 CYB\_MICIK  
38 45 31.0 430 1 RUMA\_SALTI  
39 45 31.0 430 1 RUMA\_SALTY  
40 45 31.0 432 1 RUMA\_ECOL6  
41 45 31.0 432 1 RUMA\_ECOLI  
42 45 31.0 432 1 RUMA\_ECOLI  
43 45 31.0 688 1 ST12\_YEAST  
44 45 31.0 2291 1 SPCB\_DROME  
45 45 31.0 2301 1 POLG\_TMEVD

P29740 photobacter  
P09357 streptococc  
P19538 drosophila  
Q9MIK2 micropechis  
Q8Z446 salmonella  
Q8ZNE1 salmonella  
Q8XED8 escherichia  
Q8FSG6 escherichia  
P55335 escherichia  
P13574 saccharomyc  
Q00963 drosophila  
P13899 t genome po



RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Fusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT to be the alpha 3 chain of type IV collagen.";  
RL J. Clin. Invest. 89:592-601(1992).  
RN [7]  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RA Ding J.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NCI domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions.";  
RL J. Biol. Chem. 269:2342-2348(1994).  
RN [9]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Ninomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q36.";  
RL FEBS Lett. 424:11-16(1998).  
RN [10]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced.";  
RL J. Biol. Chem. 268:12090-12094(1993).  
RN [11]  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemlink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,  
RA Readers S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome.";  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=3;  
CC Comment-additional isoforms seem to exist. Isoforms differ in  
CC the C-terminal part of the NCI domain;  
CC Name=1;  
CC IsoId=Q01955-1; Sequence=Displayed;  
CC Name=2; Synonyms=V;  
CC IsoId=Q01955-2; Sequence=VSP\_001170;  
CC Name=3; Synonyms=L5;  
CC IsoId=Q01955-3; Sequence=VSP\_001171;  
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are  
CC colocalized and present only in basement membranes of kidney, eye,  
CC cochlea, lung and brain.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NCI) at their C-terminus, frequent interruptions of the G-  
CC x-y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.  
CC

CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NCI domain, are conserved in all known type  
CC IV collagens.  
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
CC -!- DISEASE: Antibodies against the NCI domain of alpha3(IV) mediate  
CC the autoimmune disease Goodpasture syndrome [MIM:233450], which is  
CC characterized by hematuria and pulmonary hemorrhage.  
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive  
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder  
CC characterized by progressive glomerulonephritis, renal failure,  
CC hematuria, ocular abnormalities and deafness. The recessive form  
CC occurs equally between males and females.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X80031; CAA56335.1; --  
CC EMBL; AJ288487; CAC36101.1; JOINED.  
CC EMBL; AJ288488; CAC36101.1; JOINED.  
CC EMBL; AJ288489; CAC36101.1; JOINED.  
CC EMBL; AJ288490; CAC36101.1; JOINED.  
CC EMBL; AJ288491; CAC36101.1; JOINED.  
CC EMBL; AJ288492; CAC36101.1; JOINED.  
CC EMBL; AJ288493; CAC36101.1; JOINED.  
CC EMBL; AJ288494; CAC36101.1; JOINED.  
CC EMBL; AJ288495; CAC36101.1; JOINED.  
CC EMBL; AJ288496; CAC36101.1; JOINED.  
CC EMBL; AJ288497; CAC36101.1; JOINED.  
CC EMBL; AJ288498; CAC36101.1; JOINED.  
CC EMBL; AJ288499; CAC36101.1; JOINED.  
CC EMBL; AJ288500; CAC36101.1; JOINED.  
CC EMBL; AJ288501; CAC36101.1; JOINED.  
CC EMBL; AJ288502; CAC36101.1; JOINED.  
CC EMBL; AJ288503; CAC36101.1; JOINED.  
CC EMBL; AJ288504; CAC36101.1; JOINED.  
CC EMBL; AJ288505; CAC36101.1; JOINED.  
CC EMBL; AJ288506; CAC36101.1; JOINED.  
CC EMBL; AJ288507; CAC36101.1; JOINED.  
CC EMBL; AJ288508; CAC36101.1; JOINED.  
CC EMBL; AJ288509; CAC36101.1; JOINED.  
CC EMBL; AJ288510; CAC36101.1; JOINED.  
CC EMBL; AJ288511; CAC36101.1; JOINED.  
CC EMBL; AJ288512; CAC36101.1; JOINED.  
CC EMBL; AJ288513; CAC36101.1; JOINED.  
CC EMBL; AJ288514; CAC36101.1; JOINED.  
CC EMBL; AJ288515; CAC36101.1; JOINED.  
CC EMBL; AJ288516; CAC36101.1; JOINED.  
CC EMBL; AJ288517; CAC36101.1; JOINED.  
CC EMBL; AJ288518; CAC36101.1; JOINED.  
CC EMBL; AJ288519; CAC36101.1; JOINED.  
CC EMBL; AJ288520; CAC36101.1; JOINED.  
CC EMBL; AJ288521; CAC36101.1; JOINED.  
CC EMBL; AJ288522; CAC36101.1; JOINED.  
CC EMBL; AJ288523; CAC36101.1; JOINED.  
CC EMBL; AJ288524; CAC36101.1; JOINED.  
CC EMBL; AJ288525; CAC36101.1; JOINED.  
CC EMBL; AJ288526; CAC36101.1; JOINED.  
CC EMBL; AJ288527; CAC36101.1; JOINED.  
CC EMBL; AJ288528; CAC36101.1; JOINED.  
CC EMBL; AJ288529; CAC36101.1; JOINED.  
CC EMBL; AJ288530; CAC36101.1; JOINED.  
CC EMBL; AJ288531; CAC36101.1; JOINED.  
CC EMBL; AJ288532; CAC36101.1; JOINED.  
CC EMBL; AJ288533; CAC36101.1; JOINED.  
CC EMBL; AJ288534; CAC36101.1; JOINED.  
CC EMBL; AJ288535; CAC36101.1; JOINED.  
CC EMBL; AJ288536; CAC36101.1; JOINED.

```

Query Match      84.1%; Score 122; DB 1; Length 1670;
Best Local Similarity 92.3%; Pred. No. 6.4e-10;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27
    |||||
Db 1495 QRFTHMPFLFDNVNDVNFASRNDYS 1520
    |||||

RESULT 2
CA34_BOVIN STANDARD; PRT; 471 AA.
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RA Morrison K.E., Gernino G.G., Readers S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RL encoding the bovine alpha 3 chain of type IV collagen."
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV."
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type

```

```

IV collagens.
-1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M63139; AAA62708.1; -.
PIR; A39024; A39024.
InterPro; IPR008160; Collagen.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 4.
ProDom; PD003923; ProcollagenC4; 1.
SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TPR 1 1
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match      83.4%; Score 121; DB 1; Length 471;
Best Local Similarity 88.5%; Pred. No. 2.1e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27
    |||||
Db 296 QRFTHMPFLFDNVNDVNFASRNDYS 321
    |||||

RESULT 3
CA54_CANFA STANDARD; PRT; 754 AA.
AC Q28247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 5(IV) chain (Fragment).
GN COL4A5.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed; TISSUE=Kidney;
RX MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thorne P.S., Marrano P., Bauman R., McInnes R.R.;
RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
RT human X-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV."
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-

```

CC alpha 5(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -!- DISBAS: A defect in COL4A5 has been found to be the cause of  
 CC canine X-linked hereditary nephritis (HN), a disease similar to  
 CC that in humans (also referred to as Alport syndrome) characterized  
 CC by progressive renal failure and neurosensory deafness.  
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U07888; AAB60258.1; -;  
 CC DR PIR; A55267; A55267.  
 CC DR InterPro; IPR008161; Clg\_helix.  
 CC DR InterPro; IPR008160; Collagen.  
 CC DR InterPro; IPR001442; Procollagn4\_C.  
 CC DR Pfam; PF01413; C4; 2.  
 CC DR Pfam; PF01391; Collagen; 8.  
 CC DR ProDom; PD000007; Clg\_helix; 1.  
 CC DR ProDom; PD003923; ProcollagnC4; 1.  
 CC DR SMART; SM00111; C4; 2.  
 CC DR KEGG; Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 CC FT NON\_TER 1 1  
 CC FT DOMAIN <1 530 TRIPLE-HELICAL REGION  
 CC FT DISULFID 531 >754 NON-HELICAL REGION (NC1).  
 CC FT DISULFID 552 643 OR 640 (BY SIMILARITY).  
 CC FT DISULFID 585 640 OR 643 (BY SIMILARITY).  
 CC FT DISULFID 597 603 BY SIMILARITY.  
 CC FT DISULFID 662 ? OR 754 (BY SIMILARITY).  
 CC FT DISULFID 696 754 BY SIMILARITY.  
 CC FT DISULFID 708 714 BY SIMILARITY.  
 CC FT NON\_TER 754 754  
 CC SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;  
 CC  
 CC Query Match 73.1%; Score 106; DB 1; Length 754;  
 CC Best Local Similarity 73.1%; Pred. No. 6.1e-08;  
 CC Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 2 QRTTTFPLPDVNDVNFASRNDYS 27  
 CC DB 587 RRFSTTFPFPCNNVNFASRNDYS 612  
 CC  
 CC RESULT 4  
 CC ID CAS4\_HUMAN STANDARD; PRT; 1685 AA.  
 CC AC P29406; Q16126;  
 CC DT 01-DEC-1992 (Rel. 24, Created)  
 CC DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Collagen alpha 5(IV) chain precursor.  
 CC GN COL4A5.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165049; PubMed=8120014;  
 RT Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RT J. Biol. Chem. 269:6608-6614(1994).  
 RN [2]  
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
 RX TISSUE=Kidney;  
 RC MEDLINE=92316923; PubMed=1352287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RA "Complete amino acid sequence of the human alpha 5 (IV) collagen  
 RA chain and identification of a single-base mutation in exon 23  
 RA converting glycine 521 in the collagenous domain to cysteine in an  
 RA Alport syndrome patient";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 RN [3]  
 RP SEQUENCE OF 85-1685 FROM N.A.  
 RX TISSUE=Placenta;  
 RC MEDLINE=90337990; PubMed=2380186;  
 RA Pihlajaniemi T., Pihlajaniemi E.R., Myers J.C.;  
 RA "Complete primary structure of the triple-helical region and the  
 RA carboxyl-terminal domain of a new type IV collagen chain, alpha  
 RA 5(IV).";  
 RL J. Biol. Chem. 265:13758-13766(1990).  
 RN [4]  
 RP SEQUENCE OF 924-1685 FROM N.A.  
 RX MEDLINE=91169491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RA "Characterization of the 3' half of the human type IV collagen alpha  
 RA 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 RN [5]  
 RP SEQUENCE OF 914-1685 FROM N.A.  
 RX MEDLINE=90160375; PubMed=1689491;  
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,  
 RA Tryggvason K.;  
 RA "Identification of a distinct type IV collagen alpha chain with  
 RA restricted kidney distribution and assignment of its gene to the  
 RA locus of X chromosome-linked Alport syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6]  
 RP SEQUENCE OF 1442-1471 FROM N.A.  
 RX MEDLINE=90252791; PubMed=2339699;  
 RA Myers J.C., Jones T.A., Pihlajaniemi E.R., Kadri A.S., Goddard A.D.,  
 RA Sheer D., Solomon E., Pihlajaniemi T.;  
 RA "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
 RA to the region of the X chromosome containing the Alport syndrome  
 RA locus.";  
 RL Am. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
 RA Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).  
 RX MEDLINE=94133540; PubMed=8301933;  
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
 RA Cassiman J.-J., Marynen P.;  
 RA "Differential splicing of COL4A5 mRNA in kidney and white blood  
 RA cells: a complex mutation in the COL4A5 gene of an Alport patient  
 RA deletes the NC1 domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RA "The clinical spectrum of type IV collagen mutations.";  
 RT Hum. Mutat. 9:477-499(1997).  
 RN [10]  
 RP VARIANT AS SER-1564.

RX MEDLINE=91169492; PubMed=1672282;  
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
 RA Tryggvason K.,  
 RT "Single base mutation in alpha 5 (IV) collagen chain gene converting a  
 RL conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]  
 RP VARIANT AS ARG-325.  
 RA MEDLINE=92303559; PubMed=1376965;  
 RX Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
 RA Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
 (IV) chain associated with X-linked Alport syndrome: characterization  
 RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
 RT cDNA fragments.";  
 RL Am. J. Hum. Genet. 51:135-142(1992).  
 RN [12]  
 RP VARIANT AS GLU-325.  
 RX MEDLINE=93244772; PubMed=1363780;  
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
 RA Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
 RT glutamic acid in Alport syndrome.";  
 RL Hum. Mol. Genet. 1:127-129(1992).  
 RN [13]  
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
 RX MEDLINE=94010948; PubMed=8406498;  
 RA Lemlink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
 RA Tryggvason K., Haggana-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
 RA van Oost B.A., Smets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of  
 RT patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-795;  
 RX ARG-869; ARG-872 AND CYS-1241.  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 RA Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RT common in the western United States.";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS.  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
 RA Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RT exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; ASN-664 AND  
 RP MET-1428.  
 RX MEDLINE=97094179; PubMed=8940267;  
 RA Knebelmann B., Brillat C., Forestier L., Arrondel C., Jacassier D.,  
 RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
 RA Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RT syndrome.";  
 RL Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-1498.  
 RX MEDLINE=96233932; PubMed=8829632;  
 RA Tverskaya S., Bobryna V., Tsalykova F., Ignatova M.,

RA Krasnopolskaya X., Evgrafov O.;  
 RT "Substitution of Al498D in noncollagen domain of alpha 5(IV) collagen  
 RL chain associated with adult-onset X-linked Alport syndrome.";  
 RL Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-1677.  
 RX MEDLINE=97295089; PubMed=9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RT syndrome and COL4A5 R1677Q.";  
 RL Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
 RP AND ASP-1596.  
 RX MEDLINE=98112435; PubMed=9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
 RA Trivelli A., Pinciaroli A.R., Ragaolo M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RT Alport syndrome.";  
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
 RP 802-GLY--PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
 RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.  
 RX MEDLINE=99063529; PubMed=9848783;  
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,  
 RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RT Alport syndrome using PCR and direct DNA sequencing.";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
 RX MEDLINE=20030197; PubMed=10561141;  
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,  
 RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
 RT "Detection of mutations in the COL4A5 gene in over 90% of male  
 RT patients with X-linked Alport's syndrome by RT-PCR and direct  
 RT sequencing.";  
 RL Am. J. Kidney Dis. 34:854-862(1999).  
 RN [23]  
 RP VARIANT AS ARG-822.  
 Query Match 73.1%; Score 106; DB 1; Length 1685;  
 Best Local Similarity 73.1%; Pred. No. 1.5e-07;  
 Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Cyt 2 QRTTTPFLPDNDVNDNFASRDYDS 27  
 :||:||||:|:|:|||||  
 Db 1511 RRFSTMPFWMFCNNVCFASRDYDS 1536  
 RESULT 5  
 CAL4\_HUMAN  
 ID CAL4\_HUMAN STANDARD; PRT; 1669 AA.  
 AC P02462;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89340433; PubMed=2701944;  
 RA Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;  
 RT "Structural organization of the gene for the alpha 1 chain of human  
 RT type IV collagen.";

RL J. Biol. Chem. 264:13565-13571(1989).  
RN [2]  
RP SEQUENCE OF 46-1257 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88083584; PubMed=3691802;  
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
RT "Complete primary structure of the alpha 1-chain of human basement  
membrane (type IV) collagen.";  
RL FEBS Lett. 225:188-194(1987).  
RN [3]  
RP SEQUENCE OF 1-943 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88029471; PubMed=3111751;  
RA Brazel D., Oberbauer I., Dieringer H., Babel W., Gianville R.W.,  
RT Deutzmann R., Kuehn K.;  
RT "Completion of the amino acid sequence of the alpha 1 chain of human  
basement membrane collagen (type IV) reveals 21 non-triplet  
interruptions located within the collagenous domain.";  
RL Eur. J. Biochem. 168:529-536(1987).  
RN [4]  
RP SEQUENCE OF 28-243;  
RX MEDLINE=86004708; PubMed=4043082;  
RA Gianville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
region (7S domain) of the alpha 1 (IV) chain of human basement  
membrane collagen.";  
RL Eur. J. Biochem. 152:213-219(1985).  
RN [5]  
RP SEQUENCE OF 534-1447.  
RX MEDLINE=85003629; PubMed=6434307;  
RA Babel W., Gianville R.W.;  
RT "Structure of human-basement-membrane (type IV) collagen. Complete  
amino-acid sequence of a 914-residue-long pepsin fragment from the  
alpha 1(IV) chain.";  
RL Eur. J. Biochem. 143:545-556(1984).  
RN [6]  
RP SEQUENCE OF 1256-1669 FROM N.A.  
RX MEDLINE=85207819; PubMed=2581969;  
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
RT Chong M.-C., Prockop D.J., Boyd C.D.;  
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV  
procollagen reveal an unusual homology of amino acid sequences in two  
halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687(1985).  
RN [7]  
RP SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RA Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89034231; PubMed=3182844;  
RA Soininen R., Huotari M., Hostikka S.L., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
collagen are divergently encoded on opposite DNA strands and have an  
overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220(1988).  
RN [9]  
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
carboxyterminal, non-collagenous aggregation and cross-linking domain  
of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624(1988).  
CC -1- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.

CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
alpha 6(IV), each of which can form a triple helix structure  
with 2 other chains to generate type IV collagen network.  
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the  
G-X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
CC -1- PTM: Lysines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -1- PTM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -1- PTM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC -----  
CC EMBL; M26576; AAA53098.1; JOINED.  
CC EMBL; J04217; AAA53098.1; JOINED.  
CC EMBL; M26550; AAA53098.1; JOINED.  
CC EMBL; M26540; AAA53098.1; JOINED.  
CC EMBL; M26542; AAA53098.1; JOINED.  
CC EMBL; M26543; AAA53098.1; JOINED.  
CC EMBL; M26544; AAA53098.1; JOINED.  
CC EMBL; M26545; AAA53098.1; JOINED.  
CC EMBL; M26546; AAA53098.1; JOINED.  
CC EMBL; M26547; AAA53098.1; JOINED.  
CC EMBL; M26537; AAA53098.1; JOINED.  
CC EMBL; M26538; AAA53098.1; JOINED.  
CC EMBL; M26548; AAA53098.1; JOINED.  
CC EMBL; M26549; AAA53098.1; JOINED.  
CC EMBL; M26551; AAA53098.1; JOINED.  
CC EMBL; M26552; AAA53098.1; JOINED.  
CC EMBL; M26553; AAA53098.1; JOINED.  
CC EMBL; M26554; AAA53098.1; JOINED.  
CC EMBL; M26555; AAA53098.1; JOINED.  
CC EMBL; M26556; AAA53098.1; JOINED.  
CC EMBL; M26557; AAA53098.1; JOINED.  
CC EMBL; M26539; AAA53098.1; JOINED.  
CC EMBL; M26558; AAA53098.1; JOINED.  
CC EMBL; M26559; AAA53098.1; JOINED.  
CC EMBL; M26560; AAA53098.1; JOINED.  
CC EMBL; M26561; AAA53098.1; JOINED.  
CC EMBL; M26562; AAA53098.1; JOINED.  
CC EMBL; M26536; AAA53098.1; JOINED.  
CC EMBL; M26563; AAA53098.1; JOINED.  
CC EMBL; M26541; AAA53098.1; JOINED.  
CC EMBL; M26564; AAA53098.1; JOINED.  
CC EMBL; M26565; AAA53098.1; JOINED.  
CC EMBL; M26566; AAA53098.1; JOINED.  
CC EMBL; M26567; AAA53098.1; JOINED.  
CC EMBL; M26568; AAA53098.1; JOINED.  
CC EMBL; M26569; AAA53098.1; JOINED.  
CC EMBL; M26570; AAA53098.1; JOINED.  
CC EMBL; M26571; AAA53098.1; JOINED.  
CC EMBL; M26572; AAA53098.1; JOINED.  
CC EMBL; M26573; AAA53098.1; JOINED.  
CC EMBL; M26574; AAA53098.1; JOINED.  
CC EMBL; M26575; AAA53098.1; JOINED.  
CC EMBL; Y00706; CAA69698.1; JOINED.  
CC EMBL; X03561; CAA29075.1; JOINED.  
CC EMBL; M10940; AAA52006.1; JOINED.  
CC EMBL; M13155; AAA52042.1; JOINED.  
CC EMBL; S16876; CGHU4B.  
CC Genew; HGNC:2202; COL4A1.

```
DR MIM: 120130; -
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NCI).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1662.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1662 OR 1665.
FT CONFLICT 237 238 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 K -> P (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
FT SEQUENCE 1669 AA; 160611 MW; 3EBBA5DFB9B8A84 CRC64;
Query Match 72.4%; Score 105; DB 1; Length 1669;
Best Local Similarity 73.1%; Pred. No. 2.1e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
OY 2 QRTTWPFLFDVNDVDFASRNDYS 27
DB 1495 RKFTMPFLFCNNVNCVFNDRDYS 1520
RESULT 6
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
from mouse.";
RL REBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NCI) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; J03758; AAA37439.1; -  
 DR EMBL; M23333; AAA51625.1; -  
 DR EMBL; J04594; AAA50292.1; -  
 DR EMBL; X06777; CAA29946.1; -  
 DR EMBL; X02201; CAA26132.1; -  
 DR EMBL; M15832; AAA37340.1; -  
 DR EMBL; M14042; AAA37342.1; -  
 DR EMBL; M12879; AAA37343.1; -  
 DR EMBL; M13024; -; NOT ANNOTATED CDS.  
 DR EMBL; M13025; -; NOT ANNOTATED\_CDS.  
 DR EMBL; M13026; AAA37344.1; -  
 DR EMBL; M13027; AAA37345.1; -  
 DR EMBL; M13043; AAA37346.1; -  
 DR EMBL; J04448; AAA37437.1; -  
 DR PIR; A33525; CGMS4B.  
 DR MGD; MGI-88454; Col4a1.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000007; C1g\_helix; 6.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Repeat; Hydroxylation; Connective tissue; Basement membrane;  
 KW Extracellular matrix; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT DISULFID 1441 1669 NONHELICAL REGION (NC1).  
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
 FT DISULFID 1505 1511 BY SIMILARITY.  
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).  
 FT DISULFID 1604 1665 OR 1665 (BY SIMILARITY).  
 FT DISULFID 1616 1622 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 26 26 A -> P (IN REF. 2).  
 FT CONFLICT 186 186 S -> L (IN REF. 2).  
 FT CONFLICT 319 319 Q -> S (IN REF. 2).  
 FT CONFLICT 359 359 Q -> L (IN REF. 2).  
 FT CONFLICT 403 403 L -> P (IN REF. 2).  
 FT CONFLICT 481 481 P -> L (IN REF. 2).  
 FT CONFLICT 493 493 Q -> H (IN REF. 2).  
 FT CONFLICT 712 712 S -> I (IN REF. 2).  
 FT CONFLICT 813 813 E -> Q (IN REF. 2).  
 FT CONFLICT 982 982 Q -> H (IN REF. 2).  
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).  
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 72.4%; Score 105; DB 1; Length 1669;  
 Best Local Similarity 73.1%; Pred. No. 2.1e-07;  
 Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMFLFDNDVDFNFASENDYS 27  
 Db 1495 RRPSTMPFLFCNNVNCVFNFASENDYS 1520  
 ::::|||||:|:|:|||||

RESULT 7  
 CA24\_ASCSU STANDARD; PRT; 1763 AA.  
 ID\_CA24\_ASCSU  
 AC F27393;  
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
 OC Ascariidae; Ascaris.  
 RX NCBI\_TaxID=6253;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).  
 RX MEDLINE=91340768; PubMed=1714907;  
 RA Pettitt J., Kingston I.B.;  
 RT "The complete primary structure of a nematode alpha 2(IV) collagen  
 and the partial structural organization of its gene.";  
 RL J. Biol. Chem. 266:16149-16156(1991).  
 CC -!- FUNCTION: Collagen type IV is specific for basement membranes.  
 CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
 CC Type IV collagen forms a mesh-like network linked through  
 CC intermolecular interactions between 7S domains and between NC1  
 CC domains.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=I;  
 CC IsoId=P27393-1; Sequence=Displayed;  
 CC Name=II;  
 CC IsoId=P27393-2; Sequence=VSP 001159;  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 domain (NC1) at their C-terminus, frequent interruptions of the  
 G-X-Y repeats in the long central triple-helical domain (which may  
 cause flexibility in the triple helix), and a short N-terminal  
 triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 unit (G-X-Y) are hydroxylated in some or all of the chains  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 are involved in inter- and intramolecular disulfide bonding. 12 of  
 these, located in the NC1 domain, are conserved in all known type  
 IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M67507; AA18014.1; -  
 DR PIR; S16366; S16366.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 25.  
 DR ProDom; PD000007; C1g\_helix; 6.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;  
 KW Alternative splicing; Glycoprotein; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 1763 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 27 42 7S DOMAIN.  
 FT DOMAIN 43 1529 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1530 1763 NONHELICAL REGION (NC1).  
 FT DISULFID 1548 1637 OR 1634 (BY SIMILARITY).  
 FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).  
 FT DISULFID 1593 1599 BY SIMILARITY.  
 FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).  
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).  
 FT DISULFID 1702 1709 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 249 249 O-LINKED (XYL. . .) (GLYCOSAMINOGLYCAN)  
 FT VARSPLIC 230 266 (IN ISOFORM II) (POTENTIAL).  
 FT GEQGRGPGPGPGVPSGAKTIGPGAPGMKGEK ->  
 GTDIGAGPPGPPGPPGPGTGGSIIVGPRGSGDKGVK (in







CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Eye, and Kidney;  
RX MEDLINE=94171779; PubMed=8125972;  
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.,  
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA  
RT isolation and assignment of the gene to chromosome Xq22, which is the  
RT same locus for COL4A5.";  
RL J. Biol. Chem. 269:7520-7526(1994).  
RN (2)  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=94230418; PubMed=8175748;  
RA Zhou J., Ding M., Zhao Z., Reders S.T.,  
RT "Complete primary structure of the sixth chain of human basement  
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)  
RT and comparison with five other type IV collagen chains.";  
RL J. Biol. Chem. 269:13193-13199(1994).  
RN (3)  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND  
RP LYS-1110.  
RX MEDLINE=96299642; PubMed=8661006;  
RA Zhang X., Zhou J., Reders S.T., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479(1996).  
RN (4)  
RP SEQUENCE FROM N.A.  
RA Bird C., Grahm D., Lawlor S., Wilson S.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN (5)  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=93361972; PubMed=8356449;  
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,  
RT de Paeppe A., Tryggvason K., Reders S.T.;  
RA "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RT inherited smooth muscle tumors.";  
RL Science 261:1167-1169(1993).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequence=VSP\_001174;  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D21337; BAA04809.1; -  
DR EMBL; U04845; AAA19569.2; -  
DR EMBL; U47004; AAB19038.1; -  
DR EMBL; U46959; AAB19038.1; JOINED.  
DR EMBL; U46961; AAB19038.1; JOINED.  
DR EMBL; U46962; AAB19038.1; JOINED.  
DR EMBL; U46963; AAB19038.1; JOINED.  
DR EMBL; U46964; AAB19038.1; JOINED.  
DR EMBL; U46965; AAB19038.1; JOINED.  
DR EMBL; U46966; AAB19038.1; JOINED.  
DR EMBL; U46967; AAB19038.1; JOINED.  
DR EMBL; U46968; AAB19038.1; JOINED.  
DR EMBL; U46969; AAB19038.1; JOINED.  
DR EMBL; U46970; AAB19038.1; JOINED.  
DR EMBL; U46971; AAB19038.1; JOINED.  
DR EMBL; U46972; AAB19038.1; JOINED.  
DR EMBL; U46973; AAB19038.1; JOINED.  
DR EMBL; U46974; AAB19038.1; JOINED.  
DR EMBL; U46975; AAB19038.1; JOINED.  
DR EMBL; U46976; AAB19038.1; JOINED.  
DR EMBL; U46977; AAB19038.1; JOINED.  
DR EMBL; U46978; AAB19038.1; JOINED.  
DR EMBL; U46979; AAB19038.1; JOINED.  
DR EMBL; U46980; AAB19038.1; JOINED.  
DR EMBL; U46981; AAB19038.1; JOINED.  
DR EMBL; U46982; AAB19038.1; JOINED.  
DR EMBL; U46983; AAB19038.1; JOINED.  
DR EMBL; U46984; AAB19038.1; JOINED.  
DR EMBL; U46985; AAB19038.1; JOINED.  
DR EMBL; U46986; AAB19038.1; JOINED.  
DR EMBL; U46987; AAB19038.1; JOINED.  
DR EMBL; U46988; AAB19038.1; JOINED.  
DR EMBL; U46989; AAB19038.1; JOINED.  
DR EMBL; U46990; AAB19038.1; JOINED.  
DR EMBL; U46991; AAB19038.1; JOINED.  
DR EMBL; U46992; AAB19038.1; JOINED.  
DR EMBL; U46993; AAB19038.1; JOINED.  
DR EMBL; U46994; AAB19038.1; JOINED.  
DR EMBL; U46995; AAB19038.1; JOINED.  
DR EMBL; U46996; AAB19038.1; JOINED.  
DR EMBL; U46997; AAB19038.1; JOINED.  
DR EMBL; U46998; AAB19038.1; JOINED.  
DR EMBL; U46999; AAB19038.1; JOINED.  
DR EMBL; U47000; AAB19038.1; JOINED.  
DR EMBL; U47001; AAB19038.1; JOINED.  
DR EMBL; U47002; AAB19038.1; JOINED.  
DR EMBL; U47003; AAB19038.1; JOINED.  
DR EMBL; U47004; AAB19039.1; -  
DR EMBL; U46960; AAB19039.1; JOINED.  
DR EMBL; U46961; AAB19039.1; JOINED.  
DR EMBL; U46962; AAB19039.1; JOINED.  
DR EMBL; U46963; AAB19039.1; JOINED.  
DR EMBL; U46964; AAB19039.1; JOINED.  
DR EMBL; U46965; AAB19039.1; JOINED.  
DR EMBL; U46966; AAB19039.1; JOINED.  
DR EMBL; U46967; AAB19039.1; JOINED.  
DR EMBL; U46968; AAB19039.1; JOINED.  
DR EMBL; U46969; AAB19039.1; JOINED.  
DR EMBL; U46970; AAB19039.1; JOINED.  
DR EMBL; U46971; AAB19039.1; JOINED.  
DR EMBL; U46972; AAB19039.1; JOINED.  
DR EMBL; U46973; AAB19039.1; JOINED.  
DR EMBL; U46974; AAB19039.1; JOINED.  
DR EMBL; U46975; AAB19039.1; JOINED.  
DR EMBL; U46976; AAB19039.1; JOINED.  
DR EMBL; U46977; AAB19039.1; JOINED.  
DR EMBL; U46978; AAB19039.1; JOINED.  
DR EMBL; U46979; AAB19039.1; JOINED.  
DR EMBL; U46980; AAB19039.1; JOINED.  
DR EMBL; U46981; AAB19039.1; JOINED.  
DR EMBL; U46982; AAB19039.1; JOINED.  
DR EMBL; U46983; AAB19039.1; JOINED.

```
DR EMBL; U46984; AAB19039.1; JOINED.
DR EMBL; U46985; AAB19039.1; JOINED.
DR EMBL; U46986; AAB19039.1; JOINED.
DR EMBL; U46987; AAB19039.1; JOINED.
DR EMBL; U46988; AAB19039.1; JOINED.
DR EMBL; U46989; AAB19039.1; JOINED.
DR EMBL; U46990; AAB19039.1; JOINED.
DR EMBL; U46991; AAB19039.1; JOINED.
DR EMBL; U46992; AAB19039.1; JOINED.
DR EMBL; U46993; AAB19039.1; JOINED.
DR EMBL; U46994; AAB19039.1; JOINED.
DR EMBL; U46995; AAB19039.1; JOINED.
DR EMBL; U46996; AAB19039.1; JOINED.
DR EMBL; U46997; AAB19039.1; JOINED.
DR EMBL; U46998; AAB19039.1; JOINED.
DR EMBL; U46999; AAB19039.1; JOINED.
DR EMBL; U47000; AAB19039.1; JOINED.
DR EMBL; U47001; AAB19039.1; JOINED.
DR EMBL; U47002; AAB19039.1; JOINED.
DR EMBL; U47003; AAB19039.1; JOINED.
DR EMBL; U47004; AAB19039.1; JOINED.
DR EMBL; AL034369; CAA22265.1; -.
DR EMBL; AL109943; CAB89263.1; -.
DR EMBL; AL136085; CAB86748.1; -.
DR EMBL; AL031177; CAA20120.1; -.
DR EMBL; L22763; AAB16338.1; -.
DR PIR; A54122; CGHU6B.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR GO; GO:0005203; E:extracellular matrix structural constituent; NAS.
DR GO; GO:0030198; F:extracellular matrix organization and biogen. .; NAS.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; C1g_helix; 4.
DR SMART; SM00111; C4; 2.
DR SMART; SM00111; C4; 2.
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1691 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 23 46 7S DOMAIN.

Query Match 55.9%; Score 81; DB 1; Length 1691;
Best Local Similarity 56.0%; Pred. No. 0.00081;
Matches 14; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 3 RFTTWPFLEFDVNVDFASRNDYS 27
DB 1518 RFTTWPFLEFDVNVDFASRNDYS 1542
|||||:::|::|::|::|
|::|::|::|::|::|::|::|

RESULT 11
ID CA24_MOUSE STANDARD; PRT; 1707 AA.
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; PubMed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
RA Pihlajaniemi T., Kurkinen M.;
```

```
RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen.";
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; PubMed=3011432;
RA Schwarz U., Schuppan D., Oberbaumer I., Glanville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RT chain and its comparison with the alpha 1(IV) chain.";
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; PubMed=3780963;
RA Schwarz-Magdolen U., Oberbaumer I., Kuehn K.;
RT "cDNA and protein sequence of the NCI domain of the alpha 2-chain of
RT collagen IV and its comparison with alpha 1(IV).";
RL FEBS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; PubMed=3758345;
RA Vogeli G., Horn E., Carter J., Kaytes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:29-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; PubMed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RT alpha 2(IV) collagen gene.";
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdello P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -1- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
CC -----
```



Matches 15; Conservative 3; Mismatches 6; Indels 0; Gaps 0

QY 3 RFTTPEFLDNDVDFNFASRNDYS 27  
|||||: |||||  
DB 1540 RFSTPPEFLYCNPGDVCTVASRNDKS 1564

RESULT 13

CAC44\_RABIT STANDARD; PRT; 623 AA.

P55787;  
01-NOV-1997 (Rel. 35, Created)  
01-NOV-1997 (Rel. 35, Last sequence update)  
30-MAY-2000 (Rel. 39, Last annotation update)  
Collagen alpha 4(IV) chain (Fragment).  
COLA44.  
Oryctolagus cuniculus (Rabbit).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
NCBI\_TaxID=9986;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Corneal endothelium;  
MEDLINE=93054733; PubMed=1429714;  
Kamagata Y., Mattel M.-G., Ninomiya Y.;  
Isolation and sequencing of cDNAs and genomic DNAs encoding the  
alpha 4 chain of basement membrane collagen type IV and assignment of  
the gene to the distal long arm of human chromosome 2.";  
J. Biol. Chem. 267:23753-23758 (1992).  
-!- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.  
-!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
alpha 6(IV), each of which can form a triple helix structure with  
2 other chains to generate type IV collagen network.  
-!- DOMAINS: CELLULAR LOCATION: Cell surface (Potential).  
-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the G-  
X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
-!- PM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
-!- PM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; L01477; -; NOT ANNOTATED\_CDS.  
PIR; A45137; A45137.  
InterPro; IPR008160; Collagen.  
InterPro; IPR001442; Procollagen\_4\_C.  
Pfam; PF01413; C4; 2.  
Pfam; PF01391; Collagen; 5.  
SMART; PD003923; ProcollagenC4; 1.  
SMART; SMC0111; C4; 2.  
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
NON\_TER 1 1  
DOMAIN <1 392 TRIPLE-HELICAL REGION  
DOMAIN 393 623 NONHELIICAL REGION (NC1).  
DISULFID 413 502 OR 499 (BY SIMILARITY).  
DISULFID 446 499 OR 502 (BY SIMILARITY).





```
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PDC000007; Clg_helix; 3.
DR ProDom; PDC03923; Procollagen4; 1.
DR SMART; SMO0111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT DOMAIN 39 64
FT DOMAIN 65 1459
FT DOMAIN 1460 1690
FT SITE 94 96
FT SITE 145 147
FT SITE 189 191
FT SITE 310 312
FT SITE 724 726
FT SITE 785 787
FT SITE 989 991
FT SITE 1206 1207
FT SITE 1212 1214
FT DISULFID 1480 1569
FT DISULFID 1513 1566
FT DISULFID 1525 1531
FT DISULFID 1588 1686
FT DISULFID 1622 1683
FT DISULFID 1634 1641
FT CARBOHYD 142 142
FT CARBOHYD 669 669
FT VARIANT 441 446
FT VARIANT 545 545
FT VARIANT 570 570
FT VARIANT 897 897
FT VARIANT 931 931
FT VARIANT 1004 1004
FT VARIANT 1030 1030
FT VARIANT 1201 1201
FT VARIANT 1402 1402
FT VARIANT 1572 1572
FT CONFLICT 1659 1660
FT SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64;
Query Match 44.1%; Score 64; DB 1; Length 1690;
Best Local Similarity 45.8%; Pred. No. 0.27;
Matches 11; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
QY 4 FTTMPFLFDVNDVDFASRNDYS 27
Db 1517 FSTLPFAFCNIHQVCHYQNRDS 1540
RESULT 15
ID CA14_DROME STANDARD; PRT; 1775 AA.
AC P08120;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
```

DE Collagen alpha 1(IV) chain precursor.  
GN CG25C OR DCG1.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89054012; PubMed=3142875;  
RA Blumberg B., Mackrell A.J., Fessler J.H.;  
RT "Drosophila basement membrane procollagen alpha 1(IV). II. Complete  
RT cDNA sequence, genomic structure, and general implications for  
RT supramolecular assemblies";  
RL J. Biol. Chem. 263:18328-18337(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Blumberg B.;  
RT Thesis (1987), University of California / Los Angeles, U.S.A.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Mackrell A.J.;  
RT Thesis (1992), University of California / Los Angeles, U.S.A.  
RN [4]  
RP SEQUENCE OF 1065-1775 FROM N.A.  
RX MEDLINE=87194801; PubMed=3106346;  
RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,  
RT Natzle J.E., Fessler J.H.;  
RT "Basement membrane procollagen IV and its specialized carboxyl domain  
RT are conserved in Drosophila, mouse, and human";  
RL J. Biol. Chem. 262:5947-5950(1987).  
RN [5]  
RP SEQUENCE OF 1355-1775 FROM N.A.  
RX MEDLINE=87246644; PubMed=3109906;  
RA Cecchini J.P., Knibbeher B., Mixre C., le Parco Y.;  
RT "Evidence for a type-IV-related collagen in Drosophila melanogaster.  
RT Evolutionary constancy of the carboxyl-terminal noncollagenous  
RT domain";  
RL Eur. J. Biochem. 165:587-593(1987).  
RN [6]  
RP SEQUENCE OF 762-1230 FROM N.A.  
RX MEDLINE=82197577; PubMed=6210912;  
RA Monson J.M., Natzle J., Friedman J., McCarthy B.J.;  
RT "Expression and novel structure of a collagen gene in Drosophila";  
CC Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).  
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.  
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
CC Type IV collagen forms a mesh-like network linked through  
CC intermolecular interactions between 7S domains and between NC1  
CC domains.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines are at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M23704; AAA28404.1; -;  
CC EMBL; M96575; AAA59184.1; -;  
CC EMBL; J02727; AAA28423.1; -;

```

DR EMBL; M28334; AAA28422.1; -
DR EMBL; V00200; CRA23486.2; -
DR PIR; A31893; A31893.
DR Flybase; FBgn0000299; Cg25C.
DR GO; GO:0005587; C:collagen type IV; NAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 25.
DR ProDom; PD0000007; Clg_helix; 9.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
  Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 1775 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN ? 1775 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN ? 1544 TRIPLE-HELICAL REGION.
FT DOMAIN 1545 1775 NONHELICAL REGION (NC1).
FT DISULFID 1569 1655 OR 1652 (BY SIMILARITY).
FT DISULFID 1599 1652 OR 1655 (BY SIMILARITY).
FT DISULFID 1811 1817 BY SIMILARITY.
FT DISULFID 1874 1770 OR 1767 (BY SIMILARITY).
FT DISULFID 1708 1767 OR 1770 (BY SIMILARITY).
FT DISULFID 1720 1727 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 948 948 L -> S (IN REF. 6).
FT CONFLICT 997 997 S -> T (IN REF. 6).
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).
FT CONFLICT 1373 1373 T -> I (IN REF. 5).
FT CONFLICT 1496 1496 L -> R (IN REF. 5).
FT CONFLICT 1507 1511 ETGNV -> RAGOR (IN REF. 5).
FT CONFLICT 1529 1529 E -> K (IN REF. 5).
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DESAB23149525CD CRC64;

```

```

Query Match 44.1%; Score 64; DB 1; Length 1775;
Best Local Similarity 56.5%; Pred. No. 0.29;
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 3 RFTTPELFDNVNDVDFASRND 25
   ||:|:| | | | | | | | | |
Db 1602 RFSTLPELVLSGQNNVCNYASRND 1624

```

```

Search completed: April 5, 2004, 06:59:42
Job time : 3.39952 secs

```

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:50:57 ; Search time 16.2131 seconds  
(without alignments)  
525.440 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQRTTTPFLFDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	84.1	212	6 Q28512	Q28512 macaca mula
2	122	84.1	245	4 Q9NYC4	Q9NYC4 homo sapien
3	121	83.4	203	6 Q29032	Q29032 sus scrofa
4	121	83.4	203	6 Q28682	Q28682 oryctolagus
5	121	83.4	212	6 Q28567	Q28567 ovis aries
6	116	80.0	161	11 Q61430	Q61430 mus musculus
7	116	80.0	210	6 Q28273	Q28273 canis famli
8	116	80.0	246	11 Q61435	Q61435 mus musculus
9	116	80.0	1669	11 Q9Q2S0	Q9Q2S0 mus musculus
10	112	77.2	230	11 Q63122	Q63122 rattus norv
11	106	73.1	179	11 P70165	P70165 mus musculus
12	106	73.1	253	11 Q61436	Q61436 mus musculus
13	106	73.1	585	11 Q80V57	Q80V57 mus musculus
14	106	73.1	799	11 Q8BNS7	Q8BNS7 mus musculus
15	106	73.1	886	4 Q9NUS7	Q9NUS7 homo sapien
16	106	73.1	1684	6 Q8HYC1	Q8HYC1 canis famli

17	106	73.1	1688	6 Q866Z2	Q866Z2 canis famli
18	106	73.1	1691	11 Q9BSQ2	Q9BSQ2 mus musculus
19	105	72.4	226	6 Q28271	Q28271 canis famli
20	105	72.4	226	11 Q9SLQ8	Q9SLQ8 mus musculus
21	105	72.4	229	4 Q8NFB8	Q8NFB8 homo sapien
22	105	72.4	229	4 Q9NYC5	Q9NYC5 homo sapien
23	105	72.4	979	13 Q919K3	Q919K3 gallus gall
24	105	72.4	1075	4 Q86X41	Q86X41 homo sapien
25	105	72.4	1621	4 Q9H4R9	Q9H4R9 homo sapien
26	96	66.2	1747	5 Q26640	Q26640 strongyloce
27	96	66.2	1752	5 Q07265	Q07265 strongyloce
28	84	57.9	1802	5 Q17163	Q17163 brugia mala
29	81	55.9	205	6 Q28274	Q28274 canis famli
30	81	55.9	546	11 Q99K97	Q99K97 mus musculus
31	81	55.9	1600	4 Q9UEH6	Q9UEH6 homo sapien
32	81	55.9	1691	11 Q9BSQ1	Q9BSQ1 mus musculus
33	78	53.8	202	6 Q28272	Q28272 canis famli
34	78	53.8	358	11 Q91V13	Q91V13 mus musculus
35	78	53.8	673	4 Q14052	Q14052 homo sapien
36	75	51.7	1723	5 Q9GQ81	Q9GQ81 hydra atten
37	65	44.8	1761	5 Q18407	Q18407 drosophila
38	65	44.8	1940	5 Q9VMV5	Q9VMV5 drosophila
39	64	44.1	312	11 Q64457	Q64457 mus musculus
40	64	44.1	1682	11 Q9QZ89	Q9QZ89 mus musculus
41	64	44.1	1779	5 Q9VMV4	Q9VMV4 drosophila
42	63	43.4	208	6 Q29468	Q29468 canis famli
43	63	43.4	1024	5 Q8T7S4	Q8T7S4 anopheles g
44	60	41.4	713	5 Q9GV24	Q9GV24 sarcophaga
45	59	40.7	371	8 Q8SGX2	Q8SGX2 calanaria p

## ALIGNMENTS

RESULT 1

Q28512 PRELIMINARY; PRT; 212 AA.

AC Q28512; 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).  
 GN COL4A3.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 CC NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney cortex;  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
 RA Mason P.J., Pusey C.D.;  
 RT "Properties and sequences of the Goodpasture antigen of different  
 RT mammals";  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L47280; AA91861.1; -  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001442; Procollag4\_C.  
 DR InterPro; IPR00504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; Procollag4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 DR Collagen.  
 FT NON\_TER 1 212  
 FT NON\_TER 212  
 SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;

Query Match 84.1%; Score 122; DB 6; Length 212;  
 Best Local Similarity 92.3%; Pred. No. 1.8e-10;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMTPELFNDVNDVNFASRNDYS 27  
|||||  
DB 37 QRTTMTPELFNDVNDVNCNFASRNDYS 62  
|||||

## RESULT 2

Q9NYC4 PRELIMINARY; PRT; 245 AA.

ID Q9NYC4  
AC Q9NYC4  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Tumstatin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.,  
RT "Distinct anti-tumor properties of a type IV collagen domain derived  
RT from basement membrane."  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL; AF258351; AAP72632.1;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR00504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
FT NON\_TER 1  
SQ SEQUENCE 245 AA; 26952 MW; 1BE50283549A57D CRC64;

Query Match 84.1%; Score 122; DB 4; Length 245;  
Best Local Similarity 92.3%; Pred. No. 2.1e-10;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMTPELFNDVNDVNFASRNDYS 27  
|||||

DB 70 QRTTMTPELFNDVNDVNCNFASRNDYS 95  
|||||

## RESULT 3

Q29032 PRELIMINARY; PRT; 203 AA.

ID Q29032  
AC Q29032  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
RT mammals."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; L47284; AAA91882.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR00504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR PRODOM; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1  
FT NON\_TER 203  
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 83.4%; Score 121; DB 6; Length 203;  
Best Local Similarity 88.5%; Pred. No. 2.4e-10;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMTPELFNDVNDVNFASRNDYS 27  
|||||

DB 37 QRTTMTPELFNDVNDVNCNFASRNDYS 62  
|||||

## RESULT 4

Q28682 PRELIMINARY; PRT; 203 AA.

ID Q28682  
AC Q28682  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different  
RT mammals."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; L47283; AAA91893.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR InterPro; IPR00504; RNA\_rec\_mot.  
DR Pfam; PF01413; C4; 2.  
DR SMART; SM00111; C4; 2.  
DR PRODOM; PD003923; ProcollagnC4; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.

FT NON\_TER 1  
FT NON\_TER 203  
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 83.4%; Score 121; DB 6; Length 203;  
Best Local Similarity 88.5%; Pred. No. 2.4e-10;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMTPELFNDVNDVNFASRNDYS 27  
|||||

DB 37 QRTTMTPELFNDVNDVNCNFASRNDYS 62  
|||||

## RESULT 5

Q28567 PRELIMINARY; PRT; 212 AA.

ID Q28567  
AC Q28567  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).

```
GN COL4A3
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBSJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 83.4%; Score 121; DB 6; Length 212;
Best Local Similarity 88.5%; Pred. No. 2.6e-10;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFDVNDVDFASRDY 27
DQ ||||| ||||| :||| |||||
DB 37 QRFTMPFLFCNINNVCFASRDY 62

RESULT 6
ID Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbauer I.;
RT "Cloning of the NC1 domains to the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBSJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFE8236CS CRC64;
```

```
Query Match 80.0%; Score 116; DB 11; Length 161;
Best Local Similarity 84.6%; Pred. No. 1.1e-09;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFDVNDVDFASRDY 27
DQ ||||| ||||| :||| |||||
DB 4 QRFTMPFLFCNINNVCFASRDY 29

RESULT 7
ID Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A3 gene mutation.";
RL J. Biol. Chem. 271:13821-13828 (1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; ProcollagN4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagN4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119E4CA823633D CRC64;

Query Match 80.0%; Score 116; DB 6; Length 210;
Best Local Similarity 84.6%; Pred. No. 1.5e-09;
Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTMPFLFDVNDVDFASRDY 27
DQ ||||| ||||| :||| |||||
DB 47 QRFTMPFLFCNINNVCFASRDY 72

RESULT 8
ID Q61435 PRELIMINARY; PRT; 246 AA.
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/C;
RA Miner J.H., Sanes J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
```

RT laminae: Sequence, distribution, association with laminins, and  
 RL developmental switches."  
 RN J. Cell Biol. 127:879-891(1994).

[2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Balb/c;  
 RA Miner J.H.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EXBL; Z35166; CAA84529.1; -;  
 DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR001442; Procollagn4\_C.

DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 FT NON\_TER 1

SQ SEQUENCE -246 AA; 26993 MW; A9B5434P5836F324 CRC64;  
 Query Match 80.0%; Score 116; DB 11; Length 246;  
 Best Local Similarity 84.6%; Pred. No. 1.7e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27  
 DB 71 QRTTTPFLFCNNVNCVFASRNDYS 96

RESULT 9  
 Q9QZS0 PRELIMINARY; PRT; 1669 AA.

AC Q9QZS0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha 3 collagen IV.  
 GN COL4A3.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;

EX MEDLINE=20005934; PubMed=10534397;  
 RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,

RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;  
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a

RL mouse model of alport syndrome.";  
 RL Genomics 61:113-124(1999).

DR EMBL; AF169387; AAD50449.1; -;  
 DR PIR; I48302; I48302.

DR MGD; MGI:104688; Col4a3.  
 DR GO; GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagn4\_C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.

DR ProDom; PD000007; C1g\_helix; 6.  
 DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.

KW Collagen.  
 SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 80.0%; Score 116; DB 11; Length 1669;  
 Best Local Similarity 84.6%; Pred. No. 1.4e-08;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27  
 DB 1494 QRTTTPFLFCNNVNCVFASRNDYS 1519

RESULT 10  
 Q93122 PRELIMINARY; PRT; 230 AA.

ID Q93122;  
 AC Q93122;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Alpha-3 type IV collagen (Fragment).

GN COL4A3.  
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

ON NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;  
 RX MEDLINE=98210005; PubMed=950634;

RA Ryan J.J., Katbanna I., Mason P.J., Pusey C.D., Turner A.N.;  
 RT "Sequence analysis of the 'Goodpasture antigen' of mammals."

RL Nephrol. Dial. Transplant. 13:602-607(1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;

RA Turner N.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; I47281; AAS72338.2; -;  
 DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR001442; Procollagn4\_C.  
 DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; ProcollagnC4; 1.

DR SMART; SM00111; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.

KW Collagen.  
 FT NON\_TER 230

FT NON\_TER 230  
 SQ SEQUENCE 230 AA; 25398 MW; 29549E2514CC056 CRC64;

Query Match 77.2%; Score 112; DB 11; Length 230;  
 Best Local Similarity 84.6%; Pred. No. 6.6e-09;

Matches 22; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27  
 DB 55 QRTTTPFLFCNNVNCVFASRNDYS 80

RESULT 11  
 P70165 PRELIMINARY; PRT; 179 AA.

ID P70165;  
 AC P70165;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Collagen type IV alpha5 chain (Fragment).

GN COL4A5.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.

RC STRAIN=129;  
 RT Oberbauer I.;

RT "Cloning of the NCI domains of the minor collagen IV chains of mouse

via PCR (RACE) reveals the presence of the mRNAs for alpha3(IV) and





```
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK080682; BAC37980.1; -.
DR MGI; 88456; Col14a5.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 73.1%; Score 106; DB 11; Length 799;
Best Local Similarity 73.1%; Pred. No. 2.4e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
DB 625 RRFSTMPFNFNCNNVNCNFASRNDYS 650

RESULT 15
Q9NUB7 PRELIMINARY; PRT; 886 AA.
AC Q9NUB7; 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE DA24A23.1 (Collagen, type IV, alpha 5 (Alport syndrome))
DE (Fragment).
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 73.1%; Score 106; DB 4; Length 886;
Best Local Similarity 73.1%; Pred. No. 2.4e-07;
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRETTMPFLFDNVNDVNFASRNDYS 27
DB 712 RRFSTMPFNFNCNNVNCNFASRNDYS 737

Search completed: April 5, 2004, 07:03:59
Job time : 17.2131 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:44:22 ; Search time 24.3196 Seconds  
(without alignments)  
313.688 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTTPFLFDVNDVNFASRDYDS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1980s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	145	100.0	27	6 ADA20241	Ada20241 P2 peptid
2	133	91.7	27	6 ADA20239	Ada20239 T8-3 peptid
3	127	87.6	27	6 ADA20238	Ada20238 T8 peptid
4	122	84.1	79	5 AAU75600	AAU75600 Human typ
5	122	84.1	79	6 ADA20264	Ada20264 Human tum
6	122	84.1	88	5 AAU75608	AAU75608 Human typ
7	122	84.1	88	5 AAU75607	AAU75607 Human typ
8	122	84.1	88	6 ADA20271	Ada20271 Human tum
9	122	84.1	88	6 ADA20272	Ada20272 Human tum
10	122	84.1	124	5 AAU75594	AAU75594 Human typ
11	122	84.1	124	6 ADA20258	Ada20258 Human tum
12	122	84.1	132	5 AAU75597	AAU75597 Human typ
13	122	84.1	132	6 ADA20261	Ada20261 Human tum
14	122	84.1	191	5 AAU75596	AAU75596 Human typ
15	122	84.1	191	6 ADA20260	Ada20260 Human tum
16	122	84.1	211	3 AAU755918	AAU755918 Human Goo
17	122	84.1	211	5 ABG79208	ABG79208 Human GP
18	122	84.1	218	2 AAR79164	AAR79164 Partial s
19	122	84.1	218	2 AAY44172	Aay44172 Human typ
20	122	84.1	218	3 AAY56784	Aay56784 Human alp
21	122	84.1	218	4 AAE09484	AAE09484 Human alp
22	122	84.1	232	7 ADC17697	ADC17697 Human typ
23	122	84.1	244	5 ABG79218	ABG79218 Human typ
24	122	84.1	244	5 ABG79219	ABG79219 Human Goo
25	122	84.1	244	5 ABG79217	ABG79217 Human typ

26	122	84.1	244	5 AAU75595	AAU75595 Human typ
27	122	84.1	244	6 ADA20225	Ada20225 Human typ
28	122	84.1	245	3 AAY67942	Aay67942 Human typ
29	122	84.1	245	5 AAU75589	AAU75589 Human typ
30	122	84.1	254	5 AAU75598	AAU75598 Human typ
31	122	84.1	268	2 AAY31993	Aay31993 Type IV c
32	122	84.1	268	3 AAY97555	Aay97555 Human pro
33	122	84.1	1670	7 ADD47063	Add47063 Human alp
34	121	83.4	471	2 AAR79163	AAR79163 Partial s
35	121	83.4	471	2 AAY44171	Aay44171 Bovine ty
36	121	83.4	471	3 AAY56783	Aay56783 Bovine al
37	121	83.4	471	4 AAE09483	AAE09483 Bovine al
38	112	77.2	230	7 ADD47061	Add47061 Rat Prote
39	106	73.1	229	7 ADC17699	ADC17699 Human typ
40	106	73.1	264	2 AAY31995	Aay31995 Type IV c
41	106	73.1	264	3 AAY97557	Aay97557 Human alp
42	106	73.1	309	3 AAB54044	Aab54044 Human pan
43	106	73.1	772	2 AAR23873	AAR23873 Human alp
44	106	73.1	772	2 AAU09643	AAU09643 Human typ
45	106	73.1	1685	4 ABG04839	ABG04839 Novel hum

#### ALIGNMENTS

RESULT 1

ADA20241

ID ADA20241 standard; peptide; 27 AA.

XX ADA20241;

AC ADA20241;

XX 20-NOV-2003 (first entry)

XX P2 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

XX C-terminal globular non-collagenous domain; NCI; type IV collagen;

XX cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX cytostatic; gene therapy; P2 peptide; tumstatin; human;

XX type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Asp"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Asp"

WO2003059257-A2.

24-JUL-2003.

20-DEC-2002; 2002WO-US040938.

21-DEC-2001; 2001US-00032221.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

WPI; 2003-587256/55.

New peptide, useful for preparing a composition for inhibiting tumor

growth, angiogenic activity or protein synthesis in a mammalian tissue.

Claim 65; Page 45; 240pp; English.

This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the P2 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.  
 CC  
 CC SQ Sequence 27 AA;

Query Match 100.0%; Score 145; DB 6; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-16;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFDNVDFNFASRNDYS 27  
 |||||  
 DB 1 KQRTTTPFLFDNVDFNFASRNDYS 27  
 |||||

RESULT 2  
 ADA20239  
 ID ADA20239 standard; peptide; 27 AA.

XX ADA20239;

XX 20-NOV-2003 (first entry)

XX T8-3 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytostatic; gene therapy; T8-3 peptide; tumstatin; human;  
 XX type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.  
 XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"

FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX

PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 PS Claim 63; Page 45; 240pp; English.  
 XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the T8-3 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.  
 CC  
 CC SQ Sequence 27 AA;

Query Match 91.7%; Score 133; DB 6; Length 27;  
 Best Local Similarity 92.6%; Pred. No. 2.6e-14;  
 Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQRTTTPFLFDNVDFNFASRNDYS 27  
 |||||  
 DB 1 KQRTTTPFLFDNVDFNFASRNDYS 27  
 |||||

RESULT 3  
 ADA20238

ID ADA20238 standard; peptide; 27 AA.

XX ADA20238;

XX 20-NOV-2003 (first entry)

XX T8 peptide related to human type IV collagen alpha and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytostatic; gene therapy; T8 peptide; tumstatin; human;  
 XX type IV collagen alpha 3 chain; mutant; mutein.

XX Synthetic.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX

DR WPI; 2003-587256/55.  
XX  
PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX  
PS Claim 62; Page 45; 240pp; English.  
XX  
CC This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumor growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is the amino acid sequence of the T8 peptide of the invention,  
CC derived from the amino acid sequence of tumstatin, which in turn was  
CC derived from the amino acid sequence of human type IV collagen alpha 3  
CC chain.  
XX  
XX  
SQ Sequence 27 AA;  
Query Match 87.6%; Score 127; DB 6; Length 27;  
Best Local Similarity 92.6%; Pred. No. 2.4e-13;  
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KQRFTTMPEFLFDNVNDVNFASRNDYS 27  
DB 1 KQRFTTMPEFLFDNVNDVNFASRNDYS 27  
RESULT 4  
ADAU75600  
ID AAU75600 standard; protein; 79 AA.  
XX  
AC AAU75600;  
XX  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human type IV collagen alpha 3 chain mutant, Tum-5.  
XX  
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX  
OS Homo sapiens.  
XX  
XX WO200151523-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 08-JAN-2001; 2001WO-US0000565.  
XX  
XX 07-JAN-2000; 2000US-00479118.  
XX  
XX 04-APR-2000; 2000US-00543371.  
XX  
XX 21-JUL-2000; 2000US-00625191.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
XX Kalluri R;  
XX  
XX WPI; 2002-188037/24.  
XX  
XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and

PT treating disorders involving angiogenesis.  
XX  
XX Example 40; Page: 205pp; English.  
XX  
CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues  
CC 54-132 of Tumstatin. Note: The present sequence is not shown in the  
CC specification but is derived from the wild type human Tumstatin sequence  
CC given in figure 18A (see AAU75589)  
XX  
XX  
SQ Sequence 79 AA;  
Query Match 84.1%; Score 122; DB 5; Length 79;  
Best Local Similarity 92.3%; Pred. No. 5.9e-12;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFTTMPEFLFDNVNDVNFASRNDYS 27  
DB 17 QRFTTMPEFLFDNVNDVNFASRNDYS 42  
RESULT 5  
ADA20264  
ID ADA20264 standard; protein; 79 AA.  
XX  
AC ADA20264;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX Human tumstatin deletion protein tum-5 amino acid sequence.  
XX  
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.  
XX  
OS Homo sapiens.

XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluzi R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 26; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
XX CC with anti-angiogenic properties. The invention also relates to the DNA  
XX CC sequences which encode the novel proteins. A wide variety of diseases are  
XX CC the result of undesirable angiogenesis. The formation of new capillaries  
XX CC from pre-existing vessels is essential for tumor growth and metastasis.  
XX CC Basement membrane organization is dependent on the assembly of a type IV  
XX CC collagen network which may occur through the C-terminal globular non-  
XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX CC forms are ubiquitously exhibited in human basement membranes. In the  
XX CC present invention, cell surface receptors (in particular integrins) which  
XX CC specifically bind anti-angiogenic proteins and peptides (in particular  
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX CC sequences of the invention may be useful in gene therapy. The present  
XX CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
XX CC the invention which was derived from the amino acid sequence of the alpha  
XX CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
XX CC not appear in the specification but was created by the indexer from  
XX CC information given in the specification.  
XX SQ Sequence 79 AA;  
Query Match 84.1%; Score 122; DB 6; Length 79;  
Best Local Similarity 92.3%; Pred. No. 5.9e-12;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFITMPFLFDNVNDVDFASRNDYS 27  
DB 16 QRFITMPFLFCNVNDVDFASRNDYS 41  
RESULT 6  
ID AAU75608 standard; protein; 88 AA.  
XX AC AAU75608;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.  
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 82

FT XX /note= "Wild type Cys substituted with Ala"  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PR 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2002-188037/24.  
XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX PT treating disorders involving angiogenesis.  
XX PS Claim 41; Page 153; 205pp; English.  
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX CC domain, having one or more of the characteristics selected from: (a) the  
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX CC proliferation of endothelial cells; and (c) the ability to cause  
XX CC apoptosis of endothelial cells. Also described are the following: (1) use  
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX CC analogue or allelic variant in the preparation of a medicament for  
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX CC where the angiogenesis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; or (b) by  
XX CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX CC endothelial cell apoptosis is mediated by one or more endothelial cell  
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of  
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,  
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX CC preparation of a medicament for inhibiting angiogenesis or cell  
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX CC of a medicament for treating a proliferative disease in a vertebrate,  
XX CC where the disease is characterised by angiogenesis that is mediated by  
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and  
XX CC (5) use of integrins in the preparation of a medicament for promoting or  
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
XX CC or allelic variants are useful in the preparation of a medicament for  
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or  
XX CC one or more endothelial cell integrin subunits; or by promoting or  
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or  
XX CC one or more endothelial cell integrin subunits. The medicament is useful  
XX CC in inhibiting tumour growth and for the regression of an established  
XX CC tumour. The present sequence represents the amino acid sequence of human  
XX CC type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which  
XX CC consists of residues 5-126 of Tumstatin  
XX SQ Sequence 88 AA;  
Query Match 84.1%; Score 122; DB 5; Length 88;  
Best Local Similarity 92.3%; Pred. No. 6.7e-12;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFITMPFLFDNVNDVDFASRNDYS 27  
DB 26 QRFITMPFLFCNVNDVDFASRNDYS 51

RESULT 7  
AAU75607  
ID AAU75607 standard; protein; 88 AA.  
XX AC AAU75607;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.  
XX KW Human; type IV collagen alpha 3 chain; cytotstatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX OS Homo sapiens.  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PR 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2002-188037/24.  
XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX PS Claim 32; Page 152; 205pp; English.  
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists  
CC of residues 45-132 of Tumstatin  
XX SQ Sequence 88 AA;  
Query Match 84.1%; Score 122; DB 5; Length 88;  
Best Local Similarity 92.3%; Pred. No. 6.7e-12;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 QRFTTMEPLFDNVDNDFASRNDYS 27  
Db 26 QRFTTMEPLFCNVNDVCFASRNDYS 51  
RESULT 8  
ADA20271  
ID ADA20271 standard; protein; 88 AA.  
XX AC ADA20271;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytotstatic; gene therapy; alpha 3 chain; tumstatin; human;  
KW tumstatin 45-132.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 33; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytotstatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq

CC ID33) does not appear in the specification but was created by the indexer  
CC from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 84.1%; Score 122; DB 6; Length 88;

Best Local Similarity 92.3%; Pred. No. 6.7e-12;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27

DB 25 QRTTTPFLFCVNDVNCVFASRNDYS 50

RESULT 9

ADA20272

ID ADA20272 standard; protein; 88 AA.

XX AC ADA20272;

XX DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX cytosolic; gene therapy; alpha 3 chain; tumstatin; human;  
XX tumstatin 5-125-C-A; mutant; mutein.

XX Synthetic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 81 /note= "Wild-type Cys substituted by Ala at position 125  
XX of full-length tumstatin"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; SEQ ID NO 34; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which  
XX specifically bind anti-angiogenic proteins and peptides (in particular  
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX endothelial cells and thus may exhibit cytostatic activity. The DNA

CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of  
CC the "tumstatin" protein of the invention which was derived from the amino  
CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
CC sequence (Seq ID33) does not appear in the specification but was created  
CC by the indexer from information given in the specification.

XX SQ Sequence 88 AA;

Query Match 84.1%; Score 122; DB 6; Length 88;

Best Local Similarity 92.3%; Pred. No. 6.7e-12;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDVNDVNFASRNDYS 27

DB 25 QRTTTPFLFCVNDVNCVFASRNDYS 50

RESULT 10

AAU75594

ID AAU75594 standard; protein; 124 AA.

XX AC AAU75594;

XX DT 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tumstatin 333.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US0000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00825191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.

XX Example 33; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; or (b) by  
XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX endothelial cell apoptosis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; (2) use of  
XX an antibody or peptide that specifically binds the alpha1, alpha2,  
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX preparation of a medicament for inhibiting angiogenesis or cell  
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody



This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis. Basement membrane organisation is dependent on the assembly of a type IV collagen network which may occur through the C-terminal globular non-collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2 forms are ubiquitously exhibited in human basement membranes. In the present invention, cell surface receptors (in particular integrins) which specifically bind anti-angiogenic proteins and peptides (in particular the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV collagen) are disclosed. The proteins of the invention may inhibit tumour growth, angiogenic activity in mammalian tissue or protein synthesis in endothelial cells and thus may exhibit cytostatic activity. The DNA sequences of the invention may be useful in gene therapy. The present sequence is that of tumstatin 333, an abridged form of the "tumstatin" protein of the invention which was derived from the amino acid sequence of the alpha 3 chain of human type IV collagen. Note: this sequence (Seq ID20) does not appear in the specification but was created by the indexer from information given in the specification.

Query Match 84.1%; Score 122; DB 6; Length 124;  
Best Local Similarity 92.3%; Pred. No. 1e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFITMPFLFDNVNDVDNFASRNDYS 27  
|||  
Dd 69 QRFITMPFLFCNVNDVCFNFSRNDYS 94

AAU75597 standard; protein; 132 AA.

AC AAU75597;

08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tum-2.

xx  
05 Homo sapiens.

PN WO200151523-A2.

PD 19-JUL-2001.

XX  
FF  
08-JUN-80 17:00Z ; T000Z-CM100Z

07-JAN-2000; 2000US-00513113  
04-APR-2000; 2000US-00543371

FR 21-JUL-2000; 200005-00923131.  
XX

XX  
 PA (BETH-) BEXX  
NATURAL K,

DR WPI; 2002-188037/24.

A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and treating disorders involving angiogenesis.

PS Claim 31; Page 152; 205pp; English.

CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphabeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alphav, betav or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tum-2, which consists of residues  
CC 1-132 of Tumstatin. Note: The present sequence is not shown in the  
CC specification but is derived from the wild type human Tumstatin sequence  
CC given in figure 18A (see AAU75595)

XX SQ Sequence 132 AA;  
Query Match 84.1%; Score 122; DB 5; Length 132;  
Best Local Similarity 92.3%; Pred. No. 1.1e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFTTMBFLFDVNDVNFASRNDYS 27  
DB 70 QRFTTMBFLFCNVNDVCFASRNDYS 95  
RESULT 13  
ADA20261  
ID ADA20261 standard; protein; 132 AA.  
XX AC  
XX ADA20261;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tum-2 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCL1; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tum-2.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.

XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PT Kalluri R;  
XX WPI; 2003-587256/55.  
XX N-PSDB; ADA20224.  
XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX Claim 94; SEQ ID NO 23; 240bp; English.  
XX This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NCL) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which  
XX specifically bind anti-angiogenic proteins and peptides (in particular  
XX the alpha 1, alpha 2 and alpha 3 domains of the NCL domain of type IV  
XX collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX sequences of the invention may be useful in gene therapy. The present  
XX sequence is that of tum-2, an abridged form of the "tumstatin" protein of  
XX the invention which was derived from the amino acid sequence of the alpha  
XX 3 chain of human type IV collagen. Note: This sequence (Seq ID23) does  
XX not appear in the specification but was created by the indexer from  
XX information given in the specification.  
XX SQ Sequence 132 AA;

Query Match 84.1%; Score 122; DB 6; Length 132;  
Best Local Similarity 92.3%; Pred. No. 1.1e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 QRFTTMBFLFDVNDVNFASRNDYS 27  
DB 69 QRFTTMBFLFCNVNDVCFASRNDYS 94

RESULT 14  
AAU75596  
ID AAU75596 standard; protein; 191 AA.  
XX AC AAU75596;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.  
XX KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX KW non-Goodpasture fragment; alpha3 (IV)NCL domain; alphavbeta3 integrin;  
XX KW endothelial cell proliferation; apoptosis; Arresten;  
XX KW Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX OS Homo sapiens.  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PF 07-JAN-2000; 2000US-00479118.  
XX PF 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
XX  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX  
XX Example 32; Page; 205pp; English.  
XX  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of  
CC residues 54-244 of Tumstatin. Note: The present sequence is not shown in  
CC the specification but is derived from the wild type human Tumstatin  
XX sequence given in figure 18A (see AAU75589).  
XX  
XX Sequence 191 AA;  
XX  
XX Query Match 84.1%; Score 122; DB 5; Length 191;  
XX Best Local Similarity 92.3%; Pred. No. 1.7e-11;  
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 2 QRFTMPFLFDVNDVDFASRNDYS 27  
XX 17 QRFTMPFLFDVNDVDFASRNDYS 42  
XX  
XX RESULT 15  
XX ADA20260  
XX ID ADA20260 standard; protein; 191 AA.  
XX  
XX ADA20260;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX

DE Human tumstatin deletion protein tum-1 amino acid sequence.  
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;  
KW tumstatin N53.  
XX  
XX Homo sapiens.  
XX  
XX WO2003059257-A2.  
XX  
XX 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
XX  
XX 21-DEC-2001; 2001US-00032221.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
XX Kalluri R;  
XX  
XX WPI; 2003-587256/55.  
XX N-PSDB; ADA20224.  
XX  
XX New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX Claim 94; SEQ ID NO 22; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC "tumstatin" protein of the invention which was derived from the amino  
CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
CC sequence (Seq ID22) does not appear in the specification but was created  
CC by the indexer from information given in the specification.  
XX  
XX Sequence 191 AA;  
XX  
XX Query Match 84.1%; Score 122; DB 6; Length 191;  
XX Best Local Similarity 92.3%; Pred. No. 1.7e-11;  
XX Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 2 QRFTMPFLFDVNDVDFASRNDYS 27  
XX 16 QRFTMPFLFDVNDVDFASRNDYS 41  
XX  
XX Search completed: April 5, 2004, 06:58:32  
XX Job time : 24.3196 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:04:03 ; Search time 16.9322 Seconds  
(without alignments)  
418.737 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 KQFTTMBFLFNDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071436 seqs, 262597696 residues

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145	100.0	27	14	US-10-032-221B-42
2	133	91.7	27	14	US-10-032-221B-40
3	127	87.6	27	14	US-10-032-221B-39
4	122	84.1	79	14	US-10-032-221B-26
5	122	84.1	88	14	US-10-032-221B-33
6	122	84.1	88	14	US-10-032-221B-34
7	122	84.1	124	14	US-10-032-221B-20
8	122	84.1	132	14	US-10-032-221B-23
9	122	84.1	131	14	US-10-032-221B-22
10	122	84.1	211	14	US-10-270-877-46
11	122	84.1	211	14	US-10-270-837-46
12	122	84.1	232	14	US-10-206-699-304
13	122	84.1	244	14	US-10-206-699-304
14	106	73.1	229	14	US-10-206-699-306
15	106	73.1	309	9	US-09-925-297-496

16	105	72.4	229	14	US-10-206-699-302
17	105	72.4	229	14	US-10-032-221B-2
18	105	72.4	406	9	US-09-925-302-507
19	105	72.4	1669	15	US-10-372-683-8
20	101	69.7	22	14	US-10-206-699-266
21	101	69.7	25	14	US-10-032-221B-37
22	93	64.1	25	14	US-10-032-221B-38
23	91	62.8	22	14	US-10-206-699-265
24	91	62.8	1759	15	US-10-369-493-7032
25	89	61.4	22	14	US-10-206-699-267
26	86	59.3	20	14	US-10-206-699-259
27	86	59.3	20	14	US-10-032-221B-29
28	83	57.2	46	9	US-09-864-761-48095
29	83	57.2	1744	15	US-10-369-493-5832
30	81	55.9	142	9	US-09-864-761-38021
31	81	55.9	228	14	US-10-206-699-307
32	80	55.2	18	14	US-10-206-699-254
33	80	55.2	18	14	US-10-206-699-260
34	78	53.8	227	14	US-10-206-699-303
35	78	53.8	227	14	US-10-032-221B-6
36	78	53.8	430	9	US-09-925-302-518
37	78	53.8	459	15	US-10-331-496A-27
38	78	53.8	459	15	US-10-372-683-30
39	78	53.8	1712	10	US-09-961-403-9
40	74	51.0	18	14	US-10-206-699-259
41	74	51.0	22	14	US-10-206-699-270
42	72	49.7	18	14	US-10-206-699-261
43	71	49.0	22	14	US-10-206-699-268
44	70	48.3	18	14	US-10-206-699-253
45	70	48.3	20	14	US-10-206-699-290

## ALIGNMENTS

### RESULT 1

US-10-032-221B-42  
; Sequence 42, Application US/10032221B  
; Publication No. US2003014481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/136,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 42  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: P2 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length tatin molecule, and aspartic acid has been substituted for the cy

US-10-032-221B-42  
Query Match 100.0%; Score 145; DB 14; Length 27;



```
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match      84.1%; Score 122; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 2.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
DB      25 QRFTHMPFLFCNVNDVCNFSRNDYS 50

RESULT 6
US-10-032-221B-34
; Sequence 34, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 34
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine
; OTHER INFORMATION: has been substituted for the cysteine residue at position 125 of
; OTHER INFORMATION: the full-length Tumstatin molecule)
US-10-032-221B-34
```

## US-10-032-221B-34

```
Query Match      84.1%; Score 122; DB 14; Length 88;
Best Local Similarity 92.3%; Pred. No. 2.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
DB      25 QRFTHMPFLFCNVNDVCNFSRNDYS 50
```

## RESULT 7

```
US-10-032-221B-20
; Sequence 20, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 20
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20
```

```
Query Match      84.1%; Score 122; DB 14; Length 124;
Best Local Similarity 92.3%; Pred. No. 3.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 QRFTHMPFLFDNVNDVNFASRNDYS 27
      |||||
DB      69 QRFTHMPFLFCNVNDVCNFSRNDYS 94
```

## RESULT 8

```
US-10-032-221B-23
; Sequence 23, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
```

```

; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD1
; CURRENT APPLICATION NUMBER: US/10/270,877
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-877-46

Query Match      84.1%; Score 122; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27
   ||||| ||||| ||||| ||||| |||||
DB 69 QRFTTHMPFLFCNVNDVCNFASRNDYS 94

RESULT 11
US-10-270-837-46
; Sequence 46, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-10-270-837-46

Query Match      84.1%; Score 122; DB 14; Length 211;
Best Local Similarity 92.3%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRFTHMPFLFDNVNDVNFASRNDYS 27
   ||||| ||||| ||||| ||||| |||||
DB 69 QRFTTHMPFLFCNVNDVCNFASRNDYS 94

RESULT 12
US-10-206-699-304
; Sequence 304, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain He
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 06:52:07 ; Search time 6.3414 Seconds  
(without alignments)  
219.810 Million cell updates/sec

Title: US-10-032-221B-42

Perfect score: 145

Sequence: 1 QRFTTTPFLFDVNDVDFASRNDYS 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/RTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	84.1	211	4	US-09-512-563C-46
2	122	84.1	218	2	US-08-399-889-25
3	122	84.1	218	3	US-09-167-364-25
4	122	84.1	218	3	US-08-439-897-4
5	122	84.1	268	4	US-09-589-927-6
6	122	84.1	268	4	US-09-277-665-6
7	122	84.1	268	4	US-09-589-987-6
8	121	83.4	471	2	US-08-399-889-24
9	121	83.4	471	3	US-09-167-364-24
10	121	83.4	471	3	US-08-439-897-2
11	106	73.1	264	4	US-09-589-927-10
12	106	73.1	264	4	US-09-277-665-10
13	106	73.1	264	4	US-09-589-987-10
14	105	72.4	260	4	US-09-589-927-2
15	105	72.4	260	4	US-09-277-665-2
16	105	72.4	260	4	US-09-589-987-2
17	81	55.9	260	4	US-08-589-927-12
18	81	55.9	260	4	US-09-277-665-12
19	81	55.9	260	4	US-09-589-987-12
20	78	53.8	258	4	US-09-589-927-4
21	78	53.8	258	4	US-09-277-665-4
22	78	53.8	258	4	US-09-589-987-4
23	64	44.1	260	4	US-09-589-927-8
24	64	44.1	260	4	US-08-277-665-8
25	64	44.1	260	4	US-09-589-987-8
26	55	37.9	411	4	US-09-252-991A-23375
27	50	34.5	334	4	US-09-252-991A-29546

28 46 31.7 1464 3 US-08-891-640-2 Sequence 2, Appli  
29 46 31.7 1674 2 US-08-988-542C-12 Sequence 12, Appl  
30 46 31.7 1674 4 US-09-554-467A-12 Sequence 12, Appl  
31 45.5 31.4 165 4 US-09-540-236-2442 Sequence 2442, Ap  
32 45.5 31.4 288 2 US-08-424-641B-11 Sequence 11, Appl  
33 45.5 31.4 288 2 US-08-820-980-11 Sequence 11, Appl  
34 45.5 31.4 288 2 US-08-826-439-11 Sequence 11, Appl  
35 45 31.0 2548 4 US-09-172-422-1 Sequence 1, Appli  
36 44 30.3 90 3 US-08-338-907-129 Sequence 129, App  
37 44 30.3 90 4 US-09-218-207-129 Sequence 139, App  
38 44 30.3 135 4 US-09-328-352-5695 Sequence 5695, Ap  
39 44 30.3 165 4 US-09-134-000C-3704 Sequence 3704, Ap  
40 44 30.3 205 4 US-09-489-039A-13421 Sequence 13421, A  
41 44 30.3 334 4 US-09-543-681A-8177 Sequence 8177, Ap  
42 44 30.3 1234 4 US-09-489-039A-8741 Sequence 8741, Ap  
43 44 30.3 1694 1 US-08-494-168-2 Sequence 2, Appli  
44 43.5 30.0 491 4 US-09-362-899-3 Sequence 3, Appli  
45 43 29.7 309 4 US-09-107-532A-4994 Sequence 4994, Ap

#### ALIGNMENTS

RESULT 1  
US-09-512-563C-46  
; Sequence 46, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512.563C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-09-512-563C-46

Query Match 84.1%; Score 122; DB 4; Length 211;  
Best Local Similarity 92.3%; Pred. No. 1.2e-11;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTTTPFLFDVNDVDFASRNDYS 27  
Db 69 QRFTTTPFLFCVNDVDFASRNDYS 94

RESULT 2  
US-08-399-889-25  
; Sequence 25, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 218  
; TYPE: PRT

```
; ORGANISM: Human
US-08-399-889-25

Query Match      84.1%; Score 122; DB 2; Length 218;
Best Local Similarity 92.3%; Pred. NO. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCNFAFRNDYS 68

RESULT 3
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 9512638
; CURRENT APPLICATION NUMBER: US/09/167,364
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      84.1%; Score 122; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. NO. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCNFAFRNDYS 68

RESULT 4
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudecn, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      84.1%; Score 122; DB 3; Length 218;
Best Local Similarity 92.3%; Pred. NO. 1.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
Db 43 QRTTMPFLFCNVNDVCNFAFRNDYS 68

RESULT 5
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-6

Query Match      84.1%; Score 122; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. NO. 1.6e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCNFAFRNDYS 118

RESULT 6
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match      84.1%; Score 122; DB 4; Length 268;
Best Local Similarity 92.3%; Pred. NO. 1.6e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTMPFLFDNVNDVNFASRNDYS 27
Db 93 QRTTMPFLFCNVNDVCNFAFRNDYS 118

RESULT 7
US-09-589-987-6
; Sequence 6, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match      84.1%; Score 122; DB 4; Length 268;
```

Best Local Similarity 92.3%; Pred. No. 1.6e-11; Mismatches 0; Gaps 0; Indels 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27  
Db 93 QRFTHMPFLFCNVDVDFASNDYS 118

## RESULT 8

US-08-399-889-24  
; Sequence 24, Application US/08399889B  
; Patent No. 5973120  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263A  
; CURRENT APPLICATION NUMBER: US/08/399,889B  
; CURRENT FILING DATE: 1995-03-07  
; EARLIER APPLICATION NUMBER: 07/621091  
; EARLIER FILING DATE: 1990-11-30  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Calf  
US-08-399-889-24

Query Match 83.4%; Score 121; DB 2; Length 471;  
Best Local Similarity 88.5%; Pred. No. 4.6e-11;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy .2 QRFTHMPFLFDVNDVDFASNDYS 27  
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

## RESULT 9

US-09-167-364-24  
; Sequence 24, Application US/09167364  
; Patent No. 6007980  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T  
; APPLICANT: Morrison, Karen E  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 951263B  
; CURRENT APPLICATION NUMBER: US/09/167,364  
; CURRENT FILING DATE: 1998-10-07  
; EARLIER APPLICATION NUMBER: 08/399889  
; EARLIER FILING DATE: 1995-03-07  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Calf  
US-09-167-364-24

Query Match 83.4%; Score 121; DB 3; Length 471;  
Best Local Similarity 88.5%; Pred. No. 4.6e-11;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27  
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

## RESULT 10

US-09-439-897-2  
; Sequence 2, Application US/09439897

; Patent No. 6277558  
; GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-09-439-897-2

Query Match 83.4%; Score 121; DB 3; Length 471;  
Best Local Similarity 88.5%; Pred. No. 4.6e-11;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27  
Db 296 QRFTHMPFLFCNVDVDFASNDYS 321

## RESULT 11

US-09-589-927-10  
; Sequence 10, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-10

Query Match 73.1%; Score 106; DB 4; Length 264;  
Best Local Similarity 73.1%; Pred. No. 5.4e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRFTHMPFLFDVNDVDFASNDYS 27  
Db 90 RRFSTMPFMCNNDVDFASNDYS 115

## RESULT 12

US-09-277-665-10  
; Sequence 10, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-10

Query Match 73.1%; Score 106; DB 4; Length 264;  
Best Local Similarity 73.1%; Pred. No. 5.4e-09;

Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVVDNPFASRNDYS 27  
:|||||:|:|:|||||  
Db 90 RRFSTMPFMCINNVCNPFASRNDYS 115

## RESULT 13

US-09-589-987-10  
; Sequence 10, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-10

Query Match 73.1%; Score 106; DB 4; Length 264;  
Best Local Similarity 73.1%; Pred. No. 5.4e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVVDNPFASRNDYS 27  
:|||||:|:|:|||||  
Db 90 RRFSTMPFMCINNVCNPFASRNDYS 115

## RESULT 14

US-09-589-927-2  
; Sequence 2, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-2

Query Match 72.4%; Score 105; DB 4; Length 260;  
Best Local Similarity 73.1%; Pred. No. 7.6e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVVDNPFASRNDYS 27  
:|||||:|:|:|||||  
Db 86 RRFSTMPFMCINNVCNPFASRNDYS 111

## RESULT 15

US-09-277-665-2  
; Sequence 2, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665

; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-2

Query Match 72.4%; Score 105; DB 4; Length 260;  
Best Local Similarity 73.1%; Pred. No. 7.6e-09;  
Matches 19; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRTTTPFLFDNVVDNPFASRNDYS 27  
:|||||:|:|:|||||  
Db 86 RRFSTMPFMCINNVCNPFASRNDYS 111

Search completed: April 5, 2004, 07:07:26  
Job time : 6.3414 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:31:44 ; Search time 21 seconds  
(without alignments)

1117.654 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGDSGSPATWTRGF.....KAGELEKIISRCQVCMKRRH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :

PIR 78:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	100.0	1670	1 CGHU3B	collagen alpha 3(I)
2	112	45.9	220	2 B49736	collagen alpha 3(I)
3	61	25.0	81	2 A49736	collagen alpha 3(I)
4	39	16.0	246	2 I48302	collagen alpha 3(I)
5	39	16.0	471	2 A39024	collagen alpha 3(I)
6	38	15.6	161	2 S49488	collagen alpha 3(I)
7	36	14.8	52	2 S69113	collagen alpha 3(I)
8	22	9.0	79	2 C43928	probable collagen
9	17	7.0	253	2 I48304	collagen alpha 5(I)
10	17	7.0	258	2 B61228	collagen alpha 1(I)
11	17	7.0	754	2 A55267	collagen alpha 5(I)
12	17	7.0	1669	1 CGHU4B	collagen alpha 1(I)
13	17	7.0	1689	1 CQMS4B	collagen alpha 1(I)
14	17	7.0	1681	1 S22317	collagen alpha 5(I)
15	12	4.9	1752	2 A45407	collagen alpha 3(I)
16	11	4.5	58	2 B43928	probable collagen
17	11	4.5	1744	2 S40991	collagen alpha 1(I)
18	11	4.5	1761	2 T13990	collagen type IV a
19	10	4.1	1712	1 CGHU2B	collagen alpha 2(I)
20	10	4.1	1747	2 A54121	collagen alpha-4 c
21	9	3.7	312	2 I48303	collagen alpha 4(I)
22	9	3.7	775	2 A61228	collagen alpha 2(I)
23	9	3.7	1707	2 A33526	collagen alpha 2(I)
24	8	3.3	74	2 S77782	probable 1-phospho
25	8	3.3	281	2 C69709	superoxide dismuta
26	8	3.3	1691	1 QGHU6B	collagen alpha 6(I)
27	8	3.3	1775	2 A31893	collagen alpha 1(I)
28	7	2.9	49	2 S50999	superoxide dismuta
29	7	2.9	122	2 C72639	hypothetical prote

30	7	2.9	124	2 B42526	B3R protein - vacc
31	7	2.9	125	2 T08598	probable diol dehy
32	7	2.9	168	2 G72679	hypothetical prote
33	7	2.9	183	2 T47251	complex I protein
34	7	2.9	183	2 A36621	NADH2 dehydrogenas
35	7	2.9	187	2 B90191	conserved hypothet
36	7	2.9	200	2 A87671	cytochrome c oxida
37	7	2.9	202	2 S51097	superoxide dismuta
38	7	2.9	202	2 JC4396	superoxide dismuta
39	7	2.9	211	2 S34616	superoxide dismuta
40	7	2.9	211	2 S34616	superoxide dismuta
41	7	2.9	216	2 E82020	ABC transporter AT
42	7	2.9	216	2 E81247	cell division App-
43	7	2.9	233	1 DSBYN	superoxide dismuta
44	7	2.9	237	2 H84035	hypothetical prote
45	7	2.9	247	2 AB0520	conserved hypothet
46	7	2.9	255	2 JG0179	superoxide dismuta
47	7	2.9	257	2 H83575	3-deoxy-manno-octu
48	7	2.9	259	1 WMS528	complement factor
49	7	2.9	261	2 A34476	collagen alpha 2(I
50	7	2.9	264	2 E69897	hypothetical prote
51	7	2.9	267	2 AD1561	B. subtilis Yoar p
52	7	2.9	267	2 AE1204	B. subtilis Yoar p
53	7	2.9	343	2 A86241	hypothetical prote
54	7	2.9	354	2 C83577	hypothetical prote
55	7	2.9	370	2 S52599	hypothetical prote
56	7	2.9	409	2 S57689	hypothetical prote
57	7	2.9	419	2 S41607	atrolysin A (EC 3.
58	7	2.9	438	2 H89860	conserved hypothet
59	7	2.9	441	2 T38239	hypothetical prote
60	7	2.9	450	2 E70590	3-phosphoshikimate
61	7	2.9	451	2 T30603	perlecan homolog 2
62	7	2.9	453	2 S18804	collagen alpha 4(I
63	7	2.9	478	2 D72344	DNA polymerase III
64	7	2.9	550	1 HMI52	hemagglutinin prec
65	7	2.9	550	1 HMI53	hemagglutinin prec
66	7	2.9	550	1 HMI77	hemagglutinin prec
67	7	2.9	550	1 HMI80	hemagglutinin prec
68	7	2.9	550	1 HMI33	hemagglutinin prec
69	7	2.9	550	1 HMI89	hemagglutinin prec
70	7	2.9	550	1 HMI21	hemagglutinin prec
71	7	2.9	550	1 HMI98	hemagglutinin prec
72	7	2.9	550	1 HMI15	hemagglutinin prec
73	7	2.9	550	1 HMI86	hemagglutinin prec
74	7	2.9	550	2 JQ1153	hemagglutinin prec
75	7	2.9	550	2 JQ1154	hemagglutinin prec
76	7	2.9	550	2 JQ1155	hemagglutinin prec
77	7	2.9	550	2 JQ1156	hemagglutinin prec
78	7	2.9	565	1 HMI53	hemagglutinin prec
79	7	2.9	565	1 HMI54	hemagglutinin prec
80	7	2.9	565	1 HMI55	hemagglutinin prec
81	7	2.9	565	1 HMI56	hemagglutinin prec
82	7	2.9	565	1 HMI57	hemagglutinin prec
83	7	2.9	565	1 HMI58	hemagglutinin prec
84	7	2.9	565	1 HMI59	hemagglutinin prec
85	7	2.9	565	1 HMI59	hemagglutinin prec
86	7	2.9	565	1 HMI59	hemagglutinin prec
87	7	2.9	565	1 HMI59	hemagglutinin prec
88	7	2.9	565	2 S33703	hemagglutinin - in
89	7	2.9	566	1 HMI59	hemagglutinin prec
90	7	2.9	566	1 HMI59	hemagglutinin prec
91	7	2.9	566	1 HMI59	hemagglutinin prec
92	7	2.9	566	1 HMI59	hemagglutinin prec
93	7	2.9	566	1 HMI59	hemagglutinin prec
94	7	2.9	566	1 HMI59	hemagglutinin prec
95	7	2.9	623	2 A45137	collagen alpha 4(I
96	7	2.9	638	2 A30347	exotoxin A precurs
97	7	2.9	638	2 C83503	transcription fact
98	7	2.9	681	2 I38755	sodium-dependent p
99	7	2.9	693	2 S49228	probable membrane
100	7	2.9	715	2 AC0018	envelope polypepte
101	7	2.9	1225	2 T09395	DNA replication pr
102	7	2.9	1324	2 T13123	

103 103 2 2.9 1421 2 T00333  
104 104 2 2.9 1583 2 F8366  
105 105 2 2.9 1690 1 CGHUIB  
106 106 2 2.9 1758 2 T29350  
107 107 2 2.9 1759 2 T29351  
108 108 2 2.9 1763 2 S16366  
109 109 2 2.9 1966 2 T08991  
110 110 2 2.9 3660 1 S02041  
111 111 2 2.9 20 2 E39419  
112 112 2 2.5 27 2 T01664  
113 113 2 2.5 35 2 I39969  
114 114 2 2.5 36 2 PH1753  
115 115 2 2.5 44 2 AD1753  
116 116 2 2.5 49 2 AF0716  
117 117 2 2.5 56 2 S61509  
118 118 2 2.5 64 2 S75543  
119 119 2 2.5 65 2 S14712  
120 120 2 2.5 66 2 T06697  
121 121 2 2.5 72 2 AB3399  
122 122 2 2.5 76 2 D82844  
123 123 2 2.5 80 2 T45103  
124 124 2 2.5 86 2 AF2255  
125 125 2 2.5 90 2 AF1969  
126 126 2 2.5 92 2 E83334  
127 127 2 2.5 92 2 JH0716  
128 128 2 2.5 93 2 T01876  
129 129 2 2.5 96 1 CCMF55  
130 130 2 2.5 101 2 S30493  
131 131 2 2.5 103 2 B90251  
132 132 2 2.5 104 2 T44890  
133 133 2 2.5 105 2 E72599  
134 134 2 2.5 106 2 H86901  
135 135 2 2.5 107 2 T42275  
136 136 2 2.5 110 2 B71524  
137 137 2 2.5 110 2 T46071  
138 138 2 2.5 112 2 AF2540  
139 139 2 2.5 113 2 C31769  
140 140 2 2.5 113 2 AH1784  
141 141 2 2.5 115 2 E72548  
142 142 2 2.5 119 2 T49363  
143 143 2 2.5 119 2 S10914  
144 144 2 2.5 126 2 B95415  
145 145 2 2.5 127 2 A72388  
146 146 2 2.5 128 2 S49637  
147 147 2 2.5 134 2 A81062  
148 148 2 2.5 134 2 G81807  
149 149 2 2.5 136 2 C83908  
150 150 2 2.5 139 2 F70657  
151 151 2 2.5 140 2 C49829  
152 152 2 2.5 143 2 S58429  
153 153 2 2.5 151 1 GGICPH  
154 154 2 2.5 151 2 D72706  
155 155 2 2.5 152 2 A53051  
156 156 2 2.5 153 2 A96972  
157 157 2 2.5 154 2 E84093  
158 158 2 2.5 157 2 T49554  
159 159 2 2.5 159 2 T04297  
160 160 2 2.5 161 2 S04917  
161 161 2 2.5 166 1 RGHUIA  
162 162 2 2.5 166 1 RGHUIB  
163 163 2 2.5 166 2 S04934  
164 164 2 2.5 166 2 A45751  
165 165 2 2.5 167 2 C90888  
166 166 2 2.5 167 2 J01797  
167 167 2 2.5 167 2 F85729  
168 168 2 2.5 169 2 H70377  
169 169 2 2.5 169 2 C95919  
170 170 2 2.5 169 2 C95394  
171 171 2 2.5 170 2 F82908  
172 172 2 2.5 173 2 AD3343  
173 173 2 2.5 175 2 T45900  
174 174 2 2.5 176 2 F69179  
175 175 2 2.5 178 2 F84792

176 176 2 2.5 180 2 A98248  
177 177 2 2.5 182 2 T19126  
178 178 2 2.5 183 2 B90092  
179 179 2 2.5 185 2 D87274  
180 180 2 2.5 188 2 T25883  
181 181 2 2.5 189 1 IVBOIA  
182 182 2 2.5 190 2 E95420  
183 183 2 2.5 193 2 B82063  
184 184 2 2.5 194 2 G75328  
185 185 2 2.5 198 2 JE0228  
186 186 2 2.5 199 2 A89774  
187 187 2 2.5 201 2 G95088  
188 188 2 2.5 201 2 B97956  
189 189 2 2.5 203 1 S65033  
190 190 2 2.5 203 2 A42710  
191 191 2 2.5 206 2 A75508  
192 192 2 2.5 207 2 G84157  
193 193 2 2.5 209 2 S07725  
194 194 2 2.5 209 2 E87589  
195 195 2 2.5 211 2 D37471  
196 196 2 2.5 212 2 T49680  
197 197 2 2.5 213 2 H83682  
198 198 2 2.5 214 2 F72664  
199 199 2 2.5 215 2 AE2177  
200 200 2 2.5 217 1 C64411

## ALIGNMENTS

## RESULT 1

## CGHUIB

collagen alpha 3(IV) chain precursor, long splice form - human  
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
C:Species: Homo sapiens (man)  
C:Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
C:Accession: A54763; A43928; A44043; A45971; A39785  
R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Rees, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpressor  
A:Reference number: A54763; MUID:94364994; PMID:8083201  
A:Accession: A54763

## A:Molecule type: mRNA

A:Residues: 1-1670 <MAR>  
A:Cross-references: GB:X60031; NID:G577563; PID:G577564  
A:Experimental source: kidney  
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha 3(IV) chain long splice form  
A:Reference number: A43928; MUID:92147878; PMID:1737849  
A:Accession: A43928

## A:Molecule type: mRNA

A:Residues: 1331-1524, 'I', 1526-1670 <TUR>

## A:Cross-references: GB:M01379

A:Experimental source: kidney  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 13780-13784, 1992

## A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture

A:Reference number: A44043; MUID:93015826; PMID:1400291

## A:Accession: A44043

## A:Molecule type: DNA; mRNA

A:Residues: 1386-1670 <QUI>

A:Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896

A:Note: sequence extracted from NCBI backbone (NCBIP:115597)

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; MUID:94274734; PMID:8006044

A:Contents: annotation; erratum; correction to intronic sequence in A44043

R:Bernal, D.; Quinones, S.; Saus, J.

J. Biol. Chem. 268, 12090-12094, 1993

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

A:Reference number: A45971; MUID:93280184; PMID:8505332



A;Accession: A45971  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1427-1444 <ER>  
A;Note: sequence extracted from NCBI backbone (NCBIIP:133363); sequence incorrectly identified  
R;Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reeder, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1991  
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
A;Reference number: A39786; MUID:91353570; PMID:1882840  
A;Accession: A39786  
A;Molecule type: mRNA  
A;Residues: 1453-1593, 'A', 1595-1670 <MOR>  
A;Cross-references: GB:S55790; NID:G234418; PIDN:AAB19637.1; PID:G234419  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
A;Note: 1385/1; 1418/1; 1498/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A;Introns: the alpha 3 (IV) and alpha 4 (IV) chain genes are encoded on opposite strands with  
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
er associations in the interrupted helical domain (with disulfide and desmosine cross-li  
C;Function:  
A;Description: minor structural component of extracellular basement membrane in kidney d  
C;Superfamily: collagen alpha 1 (IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
F;1-28/Domain: signal sequence #status predicted <SIG>  
F;29-1670/Product: collagen alpha 3 (IV) chain, long splice form #status predicted <MAT>  
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>  
F;43-1438/Region: interrupted helical  
F;791-793/Region: cell attachment (R-G-D) motif  
F;996-998/Region: cell attachment (R-G-D) motif  
F;1154-1156/Region: cell attachment (R-G-D) motif  
F;1306-1308/Region: cell attachment (R-G-D) motif  
F;1345-1347/Region: cell attachment (R-G-D) motif  
F;1432-1434/Region: cell attachment (R-G-D) motif  
F;1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NC1>  
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F;31-33-39-41-125-422-476-479-582-722-809-1387/Disulfide bonds: interchain #status predi  
F;253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1460-1548-1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F;1505-1511-1616-1622/Disulfide bonds: #status predicted  
F;1570-1662-1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 244; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 2.1e-239;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60  
Db 1427 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 1486

Qy 61 LGTLGSCLORTTTPFLFCNVNDVCFASNDYSYWLSTPALPMNMAPITG 120  
Db 1487 LGTLGSCLORTTTPFLFCNVNDVCFASNDYSYWLSTPALPMNMAPITG 1546

Qy 121 RCTVCEGPAIAVHSQTTDIPCPHGWISLWKGFSFIFMTSAGSGTGQALASPGSCLE 180  
Db 1547 RCTVCEGPAIAVHSQTTDIPCPHGWISLWKGFSFIFMTSAGSGTGQALASPGSCLE 1606

Qy 181 EFRASPFLECHGRGTCNYNSVFWLASLNLPDMFRKPIPSVTKAGELEKIISRQVCM 240  
Db 1607 EFRASPFLECHGRGTCNYNSVFWLASLNLPDMFRKPIPSVTKAGELEKIISRQVCM 1666

Qy 241 KKRH 244  
Db 1667 KKRH 1670

RESULT 2  
A49736  
collagen alpha 3 (IV) chain, medium splice form - human (fragment)  
N;Contains: collagen alpha 3 (IV) chain, splice form GP-V  
C;Species: Homo sapiens (man)  
C;Date: 03-May-1994 #sequence\_revision 12-Nov-1999 #text\_change 17-Mar-2000  
C;Accession: A49736; D49736; S69111  
R;Feng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene. I  
A;Reference number: A49736; MUID:94124597; PMID:8294492  
A;Accession: A49736  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 169-220 <PEN1>  
A;Accession: A49736  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: mRNA  
A;Residues: 22-220 <PEN2>  
A;Cross-references: GDB:U02519; NID:G409106; PIDN:AAA18942.1; PID:G409107  
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank  
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wie  
Eur. J. Biochem. 229, 754-760, 1995  
A;Title: Characterization and expression of multiple alternatively spliced transcripts c  
uoc antigen and one of its alternative forms.  
A;Reference number: S69111; MUID:95278230; PMID:7758473  
A;Accession: S69111  
A;Molecule type: mRNA  
A;Residues: 1-45 169-204, 'L', 206-220 <PEN>  
C;Comment: For the complete sequence of the long splice form, see FIR:CGHU3B.  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
C;Superfamily: collagen alpha 1 (IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace  
F;1-220/Product: collagen alpha 3 (IV) chain, medium splice form (fragment) #status predi  
F;1-45,169-220/Product: collagen alpha 3 (IV) chain, splice from GP-V (fragment) #status  
F;22-220/Domain: carboxyl-terminal nonhelical, NCI <NC1>  
F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 45.9%; Score 112; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 6.4e-106;  
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 60  
Db 10 GLKGRGDSGPATWTRGTFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHQD 69

Qy 61 LGTLGSCLORTTTPFLFCNVNDVCFASNDYSYWLSTPALPMNMAPITG 112  
Db 70 LGTLGSCLORTTTPFLFCNVNDVCFASNDYSYWLSTPALPMNMAPITG 121

RESULT 3  
A49736  
collagen alpha 3 (IV) chain, short splice form - human (fragment)  
N;Contains: collagen alpha 3 (IV) chain, splice form GP-III  
C;Species: Homo sapiens (man)  
C;Date: 03-May-1994 #sequence\_revision 12-Nov-1999 #text\_change 12-Nov-1999  
C;Accession: A49736; D49736; S69112  
R;Feng, L.; Xia, Y.; Wilson, C.B.  
J. Biol. Chem. 269, 2342-2348, 1994  
A;Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene. I  
A;Reference number: A49736; MUID:94124597; PMID:8294492  
A;Accession: A49736  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 71-81 <PEN1>  
A;Accession: A49736  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: mRNA  
A;Residues: 22-81 <PEN2>

A;Cross-references: GB:U02520; NID:9408895; PIDN:AAAI8943.1.; PID:g408896  
A;Note: this is the conceptual translation of the nucleic acid submitted to GenBank  
R;Bernal, D.; Quinones, S.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A;Reference number: A45971; MUID:93280184; PMID:8505332  
A;Accession: B45971  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 71-81 <BER>  
A;Cross-references: PIDN:AA27014.1.; PID:g385563  
A;Note: sequence extracted from NCBI backbone (NCBIP:133955); sequence incorrectly identified  
R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wied  
Eur. J. Biochem. 229, 754-760, 1995  
A;Title: Characterization and expression of multiple alternatively spliced transcripts of  
autoantigen and one of its alternative forms.  
A;Reference number: S69111; MUID:95278230; PMID:7758473  
A;Accession: S69112  
A;Molecule type: mRNA  
A;Residues: 1-45, 71-81 <PEN>  
C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.  
C;Genetics:  
A;Gene: GDB:COL4A3  
A;Cross-references: GDB:128351; OMIM:120070  
A;Map position: 2q36-2q37  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
F;1-81/Product: collagen alpha 3(IV) chain, short splice form (fragment) #status predicted  
F;1-45, 71-81/Product: collagen alpha 3 (IV) chain, splice form GP-III (fragment) #status  
F;22-81/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
  
Query Match 25.0%; Score 61; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2e-54;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GLKGRDSSPATWTRGVFTRHSQTTPAIPSCPGTVPVLSGFSFLVQGNQRAHQD 60  
Db 10 GLKGRDSSPATWTRGVFTRHSQTTPAIPSCPGTVPVLSGFSFLVQGNQRAHQD 69  
  
Qy 61 L 61  
Db 70 L 70  
  
RESULT 4  
148302  
collagen alpha 3(IV) chain - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
C;Accession: 148302; S47278  
R;Miner, J.H.; Sames, J.R.  
J. Cell Biol. 127, 879-891, 1994  
A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
A;Reference number: A54979; MUID:95050957; PMID:7962065  
A;Accession: 148302  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-246 <RES>  
A;Cross-references: EMBL:Z35166; NID:9535197; PID:g535198  
C;Superfamily: collagen alpha 1(IV) chain  
  
Query Match 16.0%; Score 39; DB 2; Length 246;  
Best Local Similarity 100.0%; Pred. No. 1.2e-31;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 189 ECHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAG 227  
Db 191 ECHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAG 229  
  
RESULT 5  
A39024  
collagen alpha 3(IV) chain - bovine (fragment)

C;Species: Bos primigenius taurus (cattle)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815  
R;Morrison, K.E.; Germino, G.G.; Reeders, S.T.  
J. Biol. Chem. 266, 34-39, 1991  
A;Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the  
A;Reference number: A39024; MUID:91093146; PMID:1985905  
A;Accession: A39024  
A;Molecule type: mRNA  
A;Residues: 1-471 <MOR>  
A;Cross-references: EMBL:M63139; NID:g162886; PIDN:AAA62708.1.; PID:g162887  
R;Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane c  
A;Reference number: S18432; MUID:87222419; PMID:2438283  
A;Accession: S20672  
A;Molecule type: protein  
A;Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
R;Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
J. Biol. Chem. 263, 13374-13380, 1988  
A;Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
A;Reference number: S17802; MUID:88330844; PMID:3417661  
A;Accession: S17802  
A;Molecule type: protein  
A;Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
R;Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5465-5469, 1990  
A;Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
A;Reference number: A35167; MUID:90202779; PMID:2318822  
A;Accession: A35167  
A;Molecule type: protein  
A;Residues: 236-258 <GUN>  
R;Gunwar, S.; Ballister, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe  
J. Biol. Chem. 266, 15318-15324, 1991  
A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll  
A;Reference number: A39419; MUID:91332055; PMID:1869555  
A;Accession: C39419  
A;Molecule type: protein  
A;Residues: 236-255 <GU2>  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F;1-238/Domain: collagenous (fragment) #status predicted <COL>  
F;233-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>  
F;233-353/Domain: repeat NCI #status predicted <NCI1>  
F;354-471/Domain: repeat NCI #status predicted <NCI2>  
F;232, 238/Modified site: hydroxyproline (Pro) #status experimental  
F;306-312, 417-423/Disulfide bonds: #status predicted  
  
Query Match 16.0%; Score 39; DB 2; Length 471;  
Best Local Similarity 100.0%; Pred. No. 2.1e-31;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 107 MAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCP 145  
Db 334 MAPITGRALEPIYISRCTVCEGPAIAIAVHSQTTDIPPCP 372  
  
RESULT 6  
S49488  
collagen alpha 3(IV) chain - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 13-Aug-1999  
C;Accession: S49488  
R;Oberbauer, I.  
A;Description: Cloning of the NCI domains fo the minor collagen IV chains of mouse via f  
ells.  
A;Reference number: S49487  
A;Accession: S49488  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-161 <OBE>  
A;Cross-references: EMBL:X82205; NID:g559472; PIDN:CAA57689.1.; PID:g559916

C;Superfamily: collagen alpha 1(IV) chain

Query Match 15.6%; Score 38; DB 2; Length 161;  
Best Local Similarity 100.0%; Pred. No. 8.8e-31;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYNSFWLASLNPMPFRKPSTVKA 226  
DB 124 ECHGRGTCNYNSYNSFWLASLNPMPFRKPSTVKA 161

RESULT 7

S69113 collagen alpha 3(IV) chain, splice form GP-III/IV/V - human

C;Species: Homo sapiens (man)

C;Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 12-Nov-1999

C;Accession: S69113

R;Penades, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.; Wie

Eur. J. Biochem. 229, 754-760, 1995

A;Title: Characterization and expression of multiple alternatively spliced transcripts c

ucotagen and one of its alternative forms.

A;Reference number: S69111; MUID:95278230; PMID:7758473

A;Accession: S69113

A;Molecule type: mRNA

A;Residues: 1-52 <PEN>

C;Comment: For the complete sequence of the long splice form, see PIR:CGHU3B.

C;Superfamily: collagen alpha 1(IV) chain

Query Match 14.8%; Score 36; DB 2; Length 52;  
Best Local Similarity 100.0%; Pred. No. 3.7e-29;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGTFVTRHSQTTPSCPE 36  
DB 10 GLKGRGDSGPATWTRGTFVTRHSQTTPSCPE 45

RESULT 8

C43928

Probable collagen alpha 3(IV) chain - sheep (fragments)

C;Species: Ovis orientalis aries; Ovis ammon aries (domestic sheep)

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Mar-1999

C;Accession: C43928

R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al

A;Reference number: A43928; MUID:92147878; PMID:1737849

A;Accession: C43928

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-79 <TUR>

C;Superfamily: collagen alpha 1(IV) chain

C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 9.0%; Score 22; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 9e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYNSFWLASL 210  
DB 27 ECHGRGTCNYNSYNSFWLASL 48

RESULT 9

I48304

collagen alpha 5(IV) chain - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999

C;Accession: I48304; S47280

R;Miner, J.H.; Sanes, J.R.

J. Cell Biol. 127, 879-891, 1994

A;Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ

A;Reference number: A54979; MUID:95050957; PMID:7962065

A;Accession: I48304

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-253 <RES>

A;Cross-references: EMBL:Z35168; NID:G535201; PIDN:CAA84531.1; PID:G535202

C;Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
DB 94 VCNFASRNDYSYWLSTP 110

RESULT 10

B61228

collagen alpha 1(IV) chain - rabbit (fragment)

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 17-Mar-1999

C;Accession: B61228

R;Yamaguchi, N.; Sato, N.; Ko, J.S.; Niomiya, Y.

Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991

A;Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothe

A;Reference number: A61228; MUID:92010685; PMID:1717398

A;Accession: B61228

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-258 <YAM>

C;Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 258;  
Best Local Similarity 100.0%; Pred. No. 3e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
DB 99 VCNFASRNDYSYWLSTP 115

RESULT 11

A55267

collagen alpha 5(IV) chain - dog (fragment)

C;Species: Canis lupus familiaris (dog)

C;Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 13-Aug-1999

C;Accession: A55267

R;Zheng, K.; Thorne, P.S.; Marrano, P.; Bauml, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994

A;Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-li

en type IV.

A;Reference number: A55267; MUID:94224868; PMID:8171024

A;Accession: A55267

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-754 <ZHE>

A;Cross-references: GS:U07888; NID:G469547; PIDN:AAB60258.1; PID:G469548

C;Superfamily: collagen alpha 1(IV) chain

Query Match 7.0%; Score 17; DB 2; Length 754;  
Best Local Similarity 100.0%; Pred. No. 7.3e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
DB 602 VCNFASRNDYSYWLSTP 618

RESULT 12

CGHU4B

collagen alpha 1(IV) chain precursor - human

N;Alternate names: procollagen alpha 1(IV) chain

C;Species: Homo sapiens (man)

C>Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999  
C/Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58  
R/Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A/Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen  
A/Reference number: S16876; MUID:89340433; PMID:2701944  
A/Accession: S16876  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-1669 <SO1>  
A/Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803  
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R/Soininen, R.; Huotari, M.; Hosikaka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A/Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A/Reference number: A9690; MUID:89034231; PMID:3182844  
A/Accession: A32117  
A/Molecule type: DNA  
A/Residues: 1-28 <SO12>  
A/Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R/Poeschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A/Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane  
A/Reference number: S02738; MUID:89030632; PMID:2846280  
A/Accession: S02738  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 1-6 'L', 8-28 <POE>  
A/Cross-references: EMBL:X12784; NID:G30072  
R/Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;  
Eur. J. Biochem. 168, 529-536, 1987  
A/Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem  
A/Reference number: S00048; MUID:88029471; PMID:3311751  
A/Accession: S00048  
A/Molecule type: mRNA  
A/Residues: 1-318, 'A', 320-944 <BR1>  
A/Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A/Accession: S25826  
A/Molecule type: protein  
R/Glanville, R.W.; Qian, R.O.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1995  
A/Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (78  
A/Reference number: A23115; MUID:86004708; PMID:4043082  
A/Accession: A23115  
A/Molecule type: protein  
A/Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A/Experimental source: placenta  
A/Note: The amino end of the mature form is blocked  
R/Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1997  
A/Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty  
A/Reference number: S00207; MUID:8803584; PMID:3691802  
A/Accession: S00207  
A/Molecule type: mRNA  
A/Residues: 244-530 <SO13>  
A/Cross-references: EMBL:X00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
R/Ebke, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A/Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen  
A/Reference number: S39614; MUID:94038963; PMID:8223488  
A/Accession: S39614  
A/Molecule type: protein  
A/Residues: 371-554 <EBL>  
R/Babel, W.; Glanville, R.W.  
Eur. J. Biochem. 143, 545-556, 1984  
A/Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid se  
A/Reference number: A02863; MUID:85003629; PMID:6434307  
A/Accession: A02863  
A/Molecule type: protein  
A/Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-  
A/Experimental source: placenta  
R/Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A/Title: Pepsin fragments of human placental basement-membrane collagens showing interrui  
A/Reference number: S16908; MUID:82005835; PMID:6792033  
A/Accession: A58517  
A/Molecule type: protein  
A/Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14  
R/Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A/Title: Isolation and characterization of pepsin-solubilized human basement membrane (t  
A/Reference number: S16910; MUID:84053346; PMID:6416291  
A/Accession: S16910  
A/Molecule type: protein  
A/Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-  
A/Experimental source: placenta  
R/Philajantiemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; J  
J. Biol. Chem. 260, 7681-7687, 1985  
A/Title: cDNA clones coding for the pro-alpha-1(IV) chain of human type IV procollagen r  
A/Reference number: S01466; MUID:85207819; PMID:2581969  
A/Accession: S01466  
A/Molecule type: mRNA  
A/Residues: 1256-1669 <PIH>  
A/Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424  
R/Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kafalides, N.A.;  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A/Title: Restricted homology between human alpha-1 type IV and other procollagen chains.  
A/Reference number: S16879; MUID:85216555; PMID:2582422  
A/Accession: S16879  
A/Molecule type: mRNA  
A/Residues: 1259-1669 <BRI>  
A/Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818  
R/Oberbauer, I.; Laurant, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,  
Eur. J. Biochem. 147, 217-224, 1985  
A/Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
A/Reference number: A02864; MUID:85127033; PMID:2578961  
A/Accession: S19091  
A/Molecule type: protein  
A/Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',  
R/Siebold, B.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 176, 617-624, 1988  
A/Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm  
A/Reference number: S02550; MUID:89005112; PMID:2844531  
A/Contents: annotation; disulfide bonds  
C/Genetics:  
A/Gene: GDB:COL4A1  
A/Cross-references: GDB:119791; OMIM:120130  
A/Map position: 13q34-13q34  
A/Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/  
/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116  
C/Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(  
C/Complex: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim  
r-trimer associations in the interrupted helical domain (with disulfide and desmosine cr  
C/Function:  
A/Description: structural component of extracellular basement membrane  
C/Superfamily: collagen alpha 1(IV) chain  
C/Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication  
F1-36/Domain: signal sequence #status predicted <SIG>  
F1-27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F1-29-162/Domain: amino-terminal nonhelical, 7S <7SD>  
F1-63-1440/Domain: interrupted helical <COL>  
F1-414-452/Region: integrin binding #status experimental  
F1-597-599/Region: cell attachment (R-G-D) motif  
F1-917-919/Region: cell attachment (R-G-D) motif  
F1-968-970/Region: cell attachment (R-G-D) motif  
F1-1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F1-1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F1-1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F1-727/Modified site: blocked amino end (Ala) (in mature form) #status experimental  
F1-31-36-39, 41, 125-434, 467, 470/Disulfide bonds: interchain #status predicted  
F1-31-36-39, 41, 125-434, 467, 470/Disulfide bonds: interchain #status predicted  
F1-45-48, 78, 90, 129, 156, 172, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 61  
1081, 1084, 1099, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 134  
F1-45-48, 78, 90, 129, 156, 172, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 61  
94, 117, 132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 1340, 1356, 1371,  
F1-54, 63, 75, 84, 87, 96, 102, 105, 108, 111, 117, 120, 123, 138, 141, 147, 150, 153, 159, 167, 178, 181, 184,

419, 422, 425, 439, 445, 448, 451, 479, 485, 491, 494, 503, 512, 518, 524, 530, 546, 549, 552, 555, 561, 567  
 9, 745, 748, 751, 754, 763/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:126/Binding site: carbonylate (Asn) (covalent) #status experimental  
 F:126/Modified site: alanyline (Lys) #status predicted  
 F:172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical  
 F:272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical  
 F:446-447/Cleavage site: Gly-Ile (Gelatinase B) #status predicted  
 F:766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899,90  
 23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256,  
 431,1437/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F:1120,1268/Binding site: carbohydrate (Lys) (covalent) (partial) #status experimental  
 F:1214,1424/Modified site: 3-hydroxyproline (Pro) #status absent  
 F:1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted  
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted  
 Query Match 7.0%; Score 17; DB 1; Length 1669;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 VCNFASRNDYSYWLSTP 100  
 DB 1510 VCNFASRNDYSYWLSTP 1526  
 RESULT 13  
 CGMS4B  
 collagen alpha 1(IV) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000  
 C:Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A31766; S19  
 R:Nuthukumar, G.; Blumberg, B.; Kurkinen, M.  
 J. Biol. Chem. 264, 6310-6317, 1989  
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff  
 A:Reference number: A33525; MUID:89197932; PMID:2703490  
 A:Accession: A33525  
 A:Molecule type: mRNA  
 A:Residues: 1-1669 <EMBL>  
 A:Cross-references: EMBL:J04694; NID:G556296; PIDN:AAA50292.1; PID:G556297  
 R:Wood, L.; Theriault, N.; Vogeli, G.  
 FEBS Lett. 227, 5-8, 1988  
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al  
 A:Reference number: S01454; MUID:88112221; PMID:3338568  
 A:Accession: S01454  
 A:Molecule type: mRNA  
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71  
 A:Cross-references: EMBL:X06777  
 R:Killen, P.D.; Burbello, P.; Sakurai, Y.; Yamada, Y.  
 J. Biol. Chem. 263, 8706-8709, 1988  
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
 A:Reference number: A28066; MUID:88243724; PMID:3379041  
 A:Accession: A28066  
 A:Molecule type: mRNA  
 A:Residues: 1-129 <KIT>  
 A:Cross-references: EMBL:J03758; NID:G192659; PIDN:AAA37439.1; PID:G192670  
 R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,  
 Eur. J. Biochem. 147, 217-224, 1985  
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
 A:Reference number: A02864; MUID:85127033; PMID:2578961  
 A:Accession: A02864  
 A:Molecule type: mRNA  
 A:Residues: 1276-1659 <OBE>  
 A:Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876  
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.  
 Gene 43, 301-304, 1986  
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox  
 A:Reference number: A25636; MUID:86301886; PMID:3755692  
 A:Accession: A25636  
 A:Molecule type: mRNA  
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>  
 A:Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287

A:Note: the authors translated the codon CAG for residue 1374 as Arg  
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj  
 J. Biol. Chem. 262, 8496-8499, 1987  
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
 A:Reference number: A94680; MUID:87250460; PMID:3597383  
 A:Accession: A29301  
 A:Molecule type: mRNA  
 A:Residues: 1441-1669 <KUR>  
 A:Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115  
 R:Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y.  
 J. Biol. Chem. 263, 12310-12314, 1988  
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc  
 A:Reference number: S19079; MUID:88315019; PMID:2842328  
 A:Accession: S19079  
 A:Molecule type: DNA  
 A:Residues: 1-28 <KIT>  
 A:Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G456503  
 R:Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
 J. Biol. Chem. 263, 19274-19277, 1988  
 A:Title: Head-to-head arrangement of murine type IV collagen genes.  
 A:Reference number: A92702; MUID:89066738; PMID:3198626  
 A:Accession: A32003  
 A:Molecule type: DNA  
 A:Residues: 1-28 <KAY>  
 A:Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449  
 R:Burbello, P.D.; Martin, G.R.; Yamada, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
 A:Reference number: A94220; MUID:89071759; PMID:3200851  
 A:Accession: A31766  
 A:Molecule type: DNA  
 A:Residues: 1-28 <BUR>  
 A:Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668  
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.  
 J. Biol. Chem. 261, 6654-6657, 1986  
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
 A:Reference number: S19094; MUID:86196099; PMID:3009468  
 A:Accession: S19094  
 A:Molecule type: DNA  
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>  
 A:Cross-references: EMBL:M13027  
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.  
 FEBS Lett. 115, 297-300, 1980  
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty  
 A:Reference number: S16909; MUID:80246483; PMID:6772473  
 A:Accession: S16909  
 A:Molecule type: Protein  
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123  
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.  
 Eur. J. Biochem. 123, 505-512, 1982  
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin  
 A:Reference number: A25991; MUID:82186723; PMID:6804236  
 A:Accession: A25991  
 A:Molecule type: Protein  
 A:Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 100  
 61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119,  
 A:Accession: B25991  
 A:Molecule type: Protein  
 A:Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234,  
 3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1329  
 R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
 Eur. J. Biochem. 139, 401-410, 1984  
 A:Title: Subunit structure and assembly of the globular domain of basement-membrane coll  
 A:Reference number: S17801; MUID:84132058; PMID:6698021  
 A:Accession: S17801  
 A:Molecule type: Protein  
 A:Residues: 1435-1443 <WEB>  
 C:Genetics:  
 A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3  
 A:Note: the list of introns may be incomplete  
 C:Superfamily: collagen alpha 1(IV) chain  
 C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat  
 F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
 F:28-162/Domain: 78 <7SD>  
 F:63-1440/Domain: collagenous, triple helix <COL>  
 F:597-599/Region: cell attachment (R-G-D) motif  
 F:781-783/Region: cell attachment (R-G-D) motif  
 F:917-919/Region: cell attachment (R-G-D) motif  
 F:968-970/Region: cell attachment (R-G-D) motif  
 F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
 F:1441-1552/Region: duplication  
 F:1553-1669/Region: duplication  
 F:31,36,39,41,43,46,47,470/Disulfide bonds: interchain #status predicted  
 F:126/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:971,974,977,986,989,1001,1022,1031,1037,1040,1055,1060,1063,1075,1078,1080,1090,1129,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxypro  
 F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:1505-1511,1516-1622/Disulfide bonds: #status predicted

Query Match 7.0%; Score 17; DB 1; Length 1669;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
 |||||  
 DB 1510 VCNFASRNDYSYWLSTP 1526

RESULT 14  
 S22917  
 N:Collagen alpha 5(IV) chain precursor, renal splice form - human  
 N:Alternate names: procollagen alpha 5(IV) chain  
 N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text change 21-Jul-2000  
 C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35  
 R:Zhou, J.; Hert, J.M.; Leinonen, A.; Tryggvason, K.  
 J. Biol. Chem. 267, 12475-12481, 1992  
 A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi  
 n Alport syndrome patient.  
 A:Reference number: S22917; MUID:92316923; PMID:1352287  
 A:Accession: S22917  
 A:Molecule type: mRNA  
 A:Residues: 1-967 <ZH2>  
 A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52045.1; PID:G553234  
 R:Zhou, J.; Leinonen, A.; Tryggvason, K.  
 J. Biol. Chem. 269, 6608-6614, 1994  
 A:Title: Structure of the human type IV collagen COL4A5 gene.  
 A:Reference number: A54365; MUID:94165049; PMID:8120014  
 A:Accession: A54365  
 A:Molecule type: DNA  
 A:Residues: 1-922 <ZH2>  
 A:Cross-references: GB:U04470; NID:G453378; GB:U04520; NID:G463428; PIDN:AAC27816.1; PID  
 R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggvas  
 Science 261, 1167-1169, 1993  
 A:Title: Deletion of the paired alphas(IV) and alphas(IV) collagen genes in inherited sm  
 A:Reference number: A57079; MUID:93361972; PMID:8356449  
 A:Accession: A57079  
 A:Molecule type: DNA  
 A:Residues: 1-27 <ZH4>  
 A:Cross-references: GB:Z37153; NID:G587203; PIDN:CAA85512.1; PID:G587204  
 R:Pihtajaniemi, T.; Pohjola, E.R.; Myers, J.C.  
 J. Biol. Chem. 265, 13758-13766, 1990  
 A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin  
 A:Reference number: A37122; MUID:90337990; PMID:2380186  
 A:Accession: A37122  
 A:Molecule type: mRNA  
 A:Residues: 84-439; 'GS', 442-624, 'LALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PH>  
 A:Cross-references: GB:J05558; EMBL:M58526; NID:G1314209  
 A:Note: Submitted to the EMBL Data Library, February 1991  
 A:Note: the authors translated the codon GCC for residue 115 as Val  
 R:Renieri, A.; Seri, M.; Myers, J.C.; Pihtajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma  
 Hum. Mol. Genet. 1, 127-129, 1992  
 A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in

A:Reference number: I54317; MUID:93244772; PMID:1363780  
 A:Accession: I54317  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 313-324; 'E', 326-330 <REN>  
 A:Cross-references: GB:S59334; NID:G299946; PIDN:AAD13909.1; PID:G4261609  
 R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hooyhtyae, M.; Shows, T.B.; Tryggvason, K.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
 A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidney  
 A:Reference number: A34850; MUID:90160375; PMID:1689491  
 A:Accession: A34850  
 A:Molecule type: mRNA  
 A:Residues: 914-1264, 1271-1691 <HOS>  
 A:Cross-references: EMBL:M31115; NID:G180824; PIDN:AAA52045.1; PID:G180825  
 R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
 Genomics 9, 1-9, 1991  
 A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that  
 A:Reference number: A37969; MUID:91169491; PMID:2004755  
 A:Accession: S18950  
 A:Molecule type: DNA  
 A:Residues: 924-1264, 1271-1691 <ZH3>  
 A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB  
 R:Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M  
 Kidney Int. 44, 1316-1321, 1993  
 A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex  
 A:Reference number: I56971; MUID:94133540; PMID:8301933  
 A:Accession: I56971  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1258-1276 <GUO1>  
 A:Cross-references: GB:S69168; NID:G545095; PIDN:AAC60612.1; PID:G545096  
 A:Note: kidney splice form  
 A:Accession: I76598  
 A:Molecule type: mRNA  
 A:Residues: 1284-1291, 'FPLGYLACLV', <GUO2>  
 A:Cross-references: GB:S69169; NID:G545097; PIDN:AAC60613.1; PID:G545098  
 A:Note: frameshift mutation in patient with Alport syndrome  
 R:Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; So  
 Am. J. Hum. Genet. 46, 1024-1033, 1990  
 A:Title: Molecular cloning of alphas(IV) collagen and assignment of the gene to the resi  
 A:Reference number: A35335; MUID:90252791; PMID:2339699  
 A:Accession: A35335  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1448-1477 <MYE>  
 R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yos  
 Kidney Int. 46, 1307-1314, 1994  
 A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord  
 A:Reference number: I56975; MUID:95156893; PMID:7853788  
 A:Accession: I56975  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1595-1602 <NAK>  
 A:Cross-references: GB:S75903; NID:G913882; PIDN:AAB33374.1; PID:G913883  
 A:Note: premature termination mutation from a patient with Alport syndrome; one other mu  
 R:Leimink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;  
 Genomics 17, 485-489, 1993  
 A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo  
 A:Reference number: I54188; MUID:94010948; PMID:8406498  
 A:Accession: I54188  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1604-1607, 'VHDYKVC', <LEM>  
 A:Cross-references: GB:S65767; NID:G425563; PIDN:AAD13967.1; PID:G4261667  
 A:Note: Frameshift mutation from a patient with Alport syndrome; five other mutations ar  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C:Genetics:  
 A:Gene: GDB:COL4A5; ATS  
 A:Cross-references: GDB:120596; OMIM:303630  
 A:Map position: Xq22-Xq22



A: Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1215/1; 1248/1; 1281/1; 1314/1; 1347/1; 1380/1; 1413/1; 1446/1; 1479/1; 1512/1; 1545/1; 1578/1; 1611/1; 1644/1; 1677/1; 1710/1; 1743/1; 1776/1; 1809/1; 1842/1; 1875/1; 1908/1; 1941/1; 1974/1; 2007/1; 2040/1; 2073/1; 2106/1; 2139/1; 2172/1; 2205/1; 2238/1; 2271/1; 2304/1; 2337/1; 2370/1; 2403/1; 2436/1; 2469/1; 2502/1; 2535/1; 2568/1; 2601/1; 2634/1; 2667/1; 2700/1; 2733/1; 2766/1; 2799/1; 2832/1; 2865/1; 2898/1; 2931/1; 2964/1; 2997/1; 3030/1; 3063/1; 3096/1; 3129/1; 3162/1; 3195/1; 3228/1; 3261/1; 3294/1; 3327/1; 3360/1; 3393/1; 3426/1; 3459/1; 3492/1; 3525/1; 3558/1; 3591/1; 3624/1; 3657/1; 3690/1; 3723/1; 3756/1; 3789/1; 3822/1; 3855/1; 3888/1; 3921/1; 3954/1; 3987/1; 4020/1; 4053/1; 4086/1; 4119/1; 4152/1; 4185/1; 4218/1; 4251/1; 4284/1; 4317/1; 4350/1; 4383/1; 4416/1; 4449/1; 4482/1; 4515/1; 4548/1; 4581/1; 4614/1; 4647/1; 4680/1; 4713/1; 4746/1; 4779/1; 4812/1; 4845/1; 4878/1; 4911/1; 4944/1; 4977/1; 5010/1; 5043/1; 5076/1; 5109/1; 5142/1; 5175/1; 5208/1; 5241/1; 5274/1; 5307/1; 5340/1; 5373/1; 5406/1; 5439/1; 5472/1; 5505/1; 5538/1; 5571/1; 5604/1; 5637/1; 5670/1; 5703/1; 5736/1; 5769/1; 5802/1; 5835/1; 5868/1; 5901/1; 5934/1; 5967/1; 6000/1; 6033/1; 6066/1; 6099/1; 6132/1; 6165/1; 6198/1; 6231/1; 6264/1; 6297/1; 6330/1; 6363/1; 6396/1; 6429/1; 6462/1; 6495/1; 6528/1; 6561/1; 6594/1; 6627/1; 6660/1; 6693/1; 6726/1; 6759/1; 6792/1; 6825/1; 6858/1; 6891/1; 6924/1; 6957/1; 6990/1; 7023/1; 7056/1; 7089/1; 7122/1; 7155/1; 7188/1; 7221/1; 7254/1; 7287/1; 7320/1; 7353/1; 7386/1; 7419/1; 7452/1; 7485/1; 7518/1; 7551/1; 7584/1; 7617/1; 7650/1; 7683/1; 7716/1; 7749/1; 7782/1; 7815/1; 7848/1; 7881/1; 7914/1; 7947/1; 7980/1; 8013/1; 8046/1; 8079/1; 8112/1; 8145/1; 8178/1; 8211/1; 8244/1; 8277/1; 8310/1; 8343/1; 8376/1; 8409/1; 8442/1; 8475/1; 8508/1; 8541/1; 8574/1; 8607/1; 8640/1; 8673/1; 8706/1; 8739/1; 8772/1; 8805/1; 8838/1; 8871/1; 8904/1; 8937/1; 8970/1; 9003/1; 9036/1; 9069/1; 9102/1; 9135/1; 9168/1; 9201/1; 9234/1; 9267/1; 9300/1; 9333/1; 9366/1; 9399/1; 9432/1; 9465/1; 9498/1; 9531/1; 9564/1; 9597/1; 9630/1; 9663/1; 9696/1; 9729/1; 9762/1; 9795/1; 9828/1; 9861/1; 9894/1; 9927/1; 9960/1; 9993/1; 10026/1; 10059/1; 10092/1; 10125/1; 10158/1; 10191/1; 10224/1; 10257/1; 10290/1; 10323/1; 10356/1; 10389/1; 10422/1; 10455/1; 10488/1; 10521/1; 10554/1; 10587/1; 10620/1; 10653/1; 10686/1; 10719/1; 10752/1; 10785/1; 10818/1; 10851/1; 10884/1; 10917/1; 10950/1; 10983/1; 11016/1; 11049/1; 11082/1; 11115/1; 11148/1; 11181/1; 11214/1; 11247/1; 11280/1; 11313/1; 11346/1; 11379/1; 11412/1; 11445/1; 11478/1; 11511/1; 11544/1; 11577/1; 11610/1; 11643/1; 11676/1; 11709/1; 11742/1; 11775/1; 11808/1; 11841/1; 11874/1; 11907/1; 11940/1; 11973/1; 12006/1; 12039/1; 12072/1; 12105/1; 12138/1; 12171/1; 12204/1; 12237/1; 12270/1; 12303/1; 12336/1; 12369/1; 12402/1; 12435/1; 12468/1; 12501/1; 12534/1; 12567/1; 12600/1; 12633/1; 12666/1; 12699/1; 12732/1; 12765/1; 12798/1; 12831/1; 12864/1; 12897/1; 12930/1; 12963/1; 12996/1; 13029/1; 13062/1; 13095/1; 13128/1; 13161/1; 13194/1; 13227/1; 13260/1; 13293/1; 13326/1; 13359/1; 13392/1; 13425/1; 13458/1; 13491/1; 13524/1; 13557/1; 13590/1; 13623/1; 13656/1; 13689/1; 13722/1; 13755/1; 13788/1; 13821/1; 13854/1; 13887/1; 13920/1; 13953/1; 13986/1; 14019/1; 14052/1; 14085/1; 14118/1; 14151/1; 14184/1; 14217/1; 14250/1; 14283/1; 14316/1; 14349/1; 14382/1; 14415/1; 14448/1; 14481/1; 14514/1; 14547/1; 14580/1; 14613/1; 14646/1; 14679/1; 14712/1; 14745/1; 14778/1; 14811/1; 14844/1; 14877/1; 14910/1; 14943/1; 14976/1; 15009/1; 15042/1; 15075/1; 15108/1; 15141/1; 15174/1; 15207/1; 15240/1; 15273/1; 15306/1; 15339/1; 15372/1; 15405/1; 15438/1; 15471/1; 15504/1; 15537/1; 15570/1; 15603/1; 15636/1; 15669/1; 15702/1; 15735/1; 15768/1; 15801/1; 15834/1; 15867/1; 15900/1; 15933/1; 15966/1; 16000/1; 16033/1; 16066/1; 16099/1; 16132/1; 16165/1; 16198/1; 16231/1; 16264/1; 16297/1; 16330/1; 16363/1; 16396/1; 16429/1; 16462/1; 16495/1; 16528/1; 16561/1; 16594/1; 16627/1; 16660/1; 16693/1; 16726/1; 16759/1; 16792/1; 16825/1; 16858/1; 16891/1; 16924/1; 16957/1; 16990/1; 17023/1; 17056/1; 17089/1; 17122/1; 17155/1; 17188/1; 17221/1; 17254/1; 17287/1; 17320/1; 17353/1; 17386/1; 17419/1; 17452/1; 17485/1; 17518/1; 17551/1; 17584/1; 17617/1; 17650/1; 17683/1; 17716/1; 17749/1; 17782/1; 17815/1; 17848/1; 17881/1; 17914/1; 17947/1; 17980/1; 18013/1; 18046/1; 18079/1; 18112/1; 18145/1; 18178/1; 18211/1; 18244/1; 18277/1; 18310/1; 18343/1; 18376/1; 18409/1; 18442/1; 18475/1; 18508/1; 18541/1; 18574/1; 18607/1; 18640/1; 18673/1; 18706/1; 18739/1; 18772/1; 18805/1; 18838/1; 18871/1; 18904/1; 18937/1; 18970/1; 19003/1; 19036/1; 19069/1; 19102/1; 19135/1; 19168/1; 19201/1; 19234/1; 19267/1; 19300/1; 19333/1; 19366/1; 19399/1; 19432/1; 19465/1; 19498/1; 19531/1; 19564/1; 19597/1; 19630/1; 19663/1; 19696/1; 19729/1; 19762/1; 19795/1; 19828/1; 19861/1; 19894/1; 19927/1; 19960/1; 19993/1; 20026/1; 20059/1; 20092/1; 20125/1; 20158/1; 20191/1; 20224/1; 20257/1; 20290/1; 20323/1; 20356/1; 20389/1; 20422/1; 20455/1; 20488/1; 20521/1; 20554/1; 20587/1; 20620/1; 20653/1; 20686/1; 20719/1; 20752/1; 20785/1; 20818/1; 20851/1; 20884/1; 20917/1; 20950/1; 20983/1; 21016/1; 21049/1; 21082/1; 21115/1; 21148/1; 21181/1; 21214/1; 21247/1; 21280/1; 21313/1; 21346/1; 21379/1; 21412/1; 21445/1; 21478/1; 21511/1; 21544/1; 21577/1; 21610/1; 21643/1; 21676/1; 21709/1; 21742/1; 21775/1; 21808/1; 21841/1; 21874/1; 21907/1; 21940/1; 21973/1; 22006/1; 22039/1; 22072/1; 22105/1; 22138/1; 22171/1; 22204/1; 22237/1; 22270/1; 22303/1; 22336/1; 22369/1; 22402/1; 22435/1; 22468/1; 22499/1; 22532/1; 22565/1; 22598/1; 22631/1; 22664/1; 22697/1; 22730/1; 22763/1; 22796/1; 22829/1; 22862/1; 22895/1; 22928/1; 22961/1; 22994/1; 23027/1; 23060/1; 23093/1; 23126/1; 23159/1; 23192/1; 23225/1; 23258/1; 23291/1; 23324/1; 23357/1; 23390/1; 23423/1; 23456/1; 23489/1; 23522/1; 23555/1; 23588/1; 23621/1; 23654/1; 23687/1; 23720/1; 23753/1; 23786/1; 23819/1; 23852/1; 23885/1; 23918/1; 23951/1; 23984/1; 24017/1; 24050/1; 24083/1; 24116/1; 24149/1; 24182/1; 24215/1; 24248/1; 24281/1; 24314/1; 24347/1; 24380/1; 24413/1; 24446/1; 24479/1; 24512/1; 24545/1; 24578/1; 24611/1; 24644/1; 24677/1; 24710/1; 24743/1; 24776/1; 24809/1; 24842/1; 24875/1; 24908/1; 24941/1; 24974/1; 25007/1; 25040/1; 25073/1; 25106/1; 25139/1; 25172/1; 25205/1; 25238/1; 25271/1; 25304/1; 25337/1; 25370/1; 25403/1; 25436/1; 25469/1; 25502/1; 25535/1; 25568/1; 25601/1; 25634/1; 25667/1; 25700/1; 25733/1; 25766/1; 25799/1; 25832/1; 25865/1; 25898/1; 25931/1; 25964/1; 25997/1; 26030/1; 26063/1; 26096/1; 26129/1; 26162/1; 26195/1; 26228/1; 26261/1; 26294/1; 26327/1; 26360/1; 26393/1; 26426/1; 26459/1; 26492/1; 26525/1; 26558/1; 26591/1; 26624/1; 26657/1; 26690/1; 26723/1; 26756/1; 26789/1; 26822/1; 26855/1; 26888/1; 26921/1; 26954/1; 26987/1; 27020/1; 27053/1; 27086/1; 27119/1; 27152/1; 27185/1; 27218/1; 27251/1; 27284/1; 27317/1; 27350/1; 27383/1; 27416/1; 27449/1; 27482/1; 27515/1; 27548/1; 27581/1; 27614/1; 27647/1; 27680/1; 27713/1; 27746/1; 27779/1; 27812/1; 27845/1; 27878/1; 27911/1; 27944/1; 27977/1; 28010/1; 28043/1; 28076/1; 28109/1; 28142/1; 28175/1; 28208/1; 28241/1; 28274/1; 28307/1; 28340/1; 28373/1; 28406/1; 28439/1; 28472/1; 28505/1; 28538/1; 28571/1; 28604/1; 28637/1; 28670/1; 28703/1; 28736/1; 28769/1; 28802/1; 28835/1; 28868/1; 28901/1; 28934/1; 28967/1; 29000/1; 29033/1; 29066/1; 29099/1; 29132/1; 29165/1; 29198/1; 29231/1; 29264/1; 29297/1; 29330/1; 29363/1; 29396/1; 29429/1; 29462/1; 29495/1; 29528/1; 29561/1; 29594/1; 29627/1; 29660/1; 29693/1; 29726/1; 29759/1; 29792/1; 29825/1; 29858/1; 29891/1; 29924/1; 29957/1; 29990/1; 30023/1; 30056/1; 30089/1; 30122/1; 30155/1; 30188/1; 30221/1; 30254/1; 30287/1; 30320/1; 30353/1; 30386/1; 30419/1; 30452/1; 30485/1; 30518/1; 30551/1; 30584/1; 30617/1; 30650/1; 30683/1; 30716/1; 30749/1; 30782/1; 30815/1; 30848/1; 30881/1; 30914/1; 30947/1; 30980/1; 31013/1; 31046/1; 31079/1; 31112/1; 31145/1; 31178/1; 31211/1; 31244/1; 31277/1; 31310/1; 31343/1; 31376/1; 31409/1; 31442/1; 31475/1; 31508/1; 31541/1; 31574/1; 31607/1; 31640/1; 31673/1; 31706/1; 31739/1; 31772/1; 31805/1; 31838/1; 31871/1; 31904/1; 31937/1; 31970/1; 32003/1; 32036/1; 32069/1; 32102/1; 32135/1; 32168/1; 32201/1; 32234/1; 32267/1; 32300/1; 32333/1; 32366/1; 32399/1; 32432/1; 32465/1; 32498/1; 32531/1; 32564/1; 32597/1; 32630/1; 32663/1; 32696/1; 32729/1; 32762/1; 32795/1; 32828/1; 32861/1; 32894/1; 32927/1; 32960/1; 32993/1; 33026/1; 33059/1; 33092/1; 33125/1; 33158/1; 33191/1; 33224/1; 33257/1; 33290/1; 33323/1; 33356/1; 33389/1; 33422/1; 33455/1; 33488/1; 33521/1; 33554/1; 33587/1; 33620/1; 33653/1; 33686/1; 33719/1; 33752/1; 33785/1; 33818/1; 33851/1; 33884/1; 33917/1; 33950/1; 33983/1; 34016/1; 34049/1; 34082/1; 34115/1; 34148/1; 34181/1; 34214/1; 34247/1; 34280/1; 34313/1; 34346/1; 34379/1; 34412/1; 34445/1; 34478/1; 34511/1; 34544/1; 34577/1; 34610/1; 34643/1; 34676/1; 34709/1; 34742/1; 34775/1; 34808/1; 34841/1; 34874/1; 34907/1; 34940/1; 34973/1; 35006/1; 35039/1; 35072/1; 35105/1; 35138/1; 35171/1; 35204/1; 35237/1; 35270/1; 35303/1; 35336/1; 35369/1; 35402/1; 35435/1; 35468/1; 35501/1; 35534/1; 35567/1; 35600/1; 35633/1; 35666/1; 35699/1; 35732/1; 35765/1; 35798/1; 35831/1; 35864/1; 35897/1; 35930/1; 35963/1; 35996/1; 36029/1; 36062/1; 36095/1; 36128/1; 36161/1; 36194/1; 36227/1; 36260/1; 36293/1; 36326/1; 36359/1; 36392/1; 36425/1; 36458/1; 36491/1; 36524/1; 36557/1; 36590/1; 36623/1; 36656/1; 36689/1; 36722/1; 36755/1; 36788/1; 36821/1; 36854/1; 36887/1; 36920/1; 36953/1; 36986/1; 37019/1; 37052/1; 37085/1; 37118/1; 37151/1; 37184/1; 37217/1; 37250/1; 37283/1; 37316/1; 37349/1; 37382/1; 37415/1; 37448/1; 37481/1; 37514/1; 37547/1; 37580/1; 37613/1; 37646/1; 37679/1; 37712/1; 37745/1; 37778/1; 37811/1; 37844/1; 37877/1; 37910/1; 37943/1; 37976/1; 38009/1; 38042/1; 38075/1; 38108/1; 38141/1; 38174/1; 38207/1; 38240/1; 38273/1; 38306/1; 38339/1; 38372/1; 38405/1; 38438/1; 38471/1; 38504/1; 38537/1; 38570/1; 38603/1; 38636/1; 38669/1; 38702/1; 38735/1; 38768/1; 38801/1; 38834/1; 38867/1; 38900/1; 38933/1; 38966/1; 39000/1; 39033/1; 39066/1; 39099/1; 39132/1; 39165/1; 39198/1; 39231/1; 39264/1; 39297/1; 39330/1; 39363/1; 39396/1; 39429/1; 39462/1; 39495/1; 39528/1; 39561/1; 39594/1; 39627/1; 39660/1; 39693/1; 39726/1; 39759/1; 39792/1; 39825/1; 39858/1; 39891/1; 39924/1; 39957/1; 39990/1; 40023/1; 40056/1; 40089/1; 40122/1; 40155/1; 40188/1; 40221/1; 40254/1; 40287/1; 40320/1; 40353/1; 40386/1; 40419/1; 40452/1; 40485/1; 40518/1; 40551/1; 40584/1; 40617/1; 40650/1; 40683/1; 40716/1; 40749/1; 40782/1; 40815/1; 40848/1; 40881/1; 40914/1; 40947/1; 40980/1; 41013/1; 41046/1; 41079/1; 41112/1; 41145/1; 41178/1; 41211/1; 41244/1; 41277/1; 41310/1; 41343/1; 41376/1; 41409/1; 41442/1; 41475/1; 41508/1; 41541/1; 41574/1; 41607/1; 41640/1; 41673/1; 41706/1; 41739/1; 41772/1; 41805/1; 41838/1; 41871/1; 41904/1; 41937/1; 41970/1; 42003/1; 42036/1; 42069/1; 42102/1; 42135/1; 42168/1; 42201/1; 42234/1; 42267/1; 42300/1; 42333/1; 42366/1; 42399/1; 42432/1; 42465/1; 42498/1; 42531/1; 42564/1; 42597/1; 42630/1; 42663/1; 42696/1; 42729/1; 42762/1; 42795/1; 42828/1; 42861/1; 42894/1; 42927/1; 42960/1; 42993/1; 43026/1; 43059/1; 43092/1; 43125/1; 43158/1; 43191/1; 43224/1; 43257/1; 43290/1; 43323/1; 43356/1; 43389/1; 43422/1; 43455/1; 43488/1; 43521/1; 43554/1; 43587/1; 43620/1; 43653/1; 43686/1; 43719/1; 43752/1; 43785/1; 43818/1; 43851/1; 43884/1; 43917/1; 43950/1; 43983/1; 44016/1; 44049/1; 44082/1; 44115/1; 44148/1; 44181/1; 44214/1; 44247/1; 44280/1; 44313/1; 44346/1; 44379/1; 44412/1; 44445/1; 44478/1; 44511/1; 44544/1; 44577/1; 44610/1; 44643/1; 44676/1; 44709/1; 44742/1; 44775/1; 44808/1; 44841/1; 44874/1; 44907/1; 44940/1; 44973/1; 45006/1; 45039/1; 45072/1; 45105/1; 45138/1; 45171/1; 45204/1; 45237/1; 45270/1; 45303/1; 45336/1; 45369/1; 45402/1; 45435/1; 45468/1; 45501/1; 45534/1; 45567/1; 45600/1; 45633/1; 45666/1; 45699/1; 45732/1; 45765/1; 45798/1; 45831/1; 45864/1; 45897/1; 45930/1; 45963/1; 45996/1; 46029/1; 46062/1; 46095/1; 46128/1; 46161/1; 46194/1; 46227/1; 46260/1; 46293/1; 46326/1; 46359/1; 46392/1



A;Accession: B34476  
A;Molecule type: DNA  
A;Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>  
A;Cross-references: EMBL:J05067; NID:G156255; PIDN:AAB59179.1; PID:G156256  
C;Genetics:  
A;Gene: cib-2; emb-9  
A;Map position: 3  
A;Introns: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F;43-1515/Domain: collagenous, triple helix #status predicted <COL>  
F;93-95/Region: cell attachment (R-G-D) motif  
F;1053-1055/Region: cell attachment (R-G-D) motif  
F;1396-1398/Region: cell attachment (R-G-D) motif  
F;1516-1744/Domain: carboxyl-terminal, nonhelical, NC1 #status predicted <NC1>  
F;1516-1627, 1628-1744/Region: duplication  
F;1580-1586, 1691-1697/Disulfide bonds: #status predicted

Query Match 4.5%; Score 11; DB 2; Length 1744;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLEEFRA 184  
|||  
Db 1675 SPGSCLEEFRA 1695

RESULT 18  
collagen type IV alpha 2 - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C;Accession: T13990  
R;Yasochornsrikul, S.; Davis, W.J.; Cramer, G.; Kimbrell, D.A.; Dearolf, C.R.  
submitted to the EMBL Data Library, July 1996  
A;Description: Viking: identification and characterization of a novel type IV collagen  
A;Reference number: Z17845  
A;Accession: T13990  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1761 <YAS>  
A;Cross-references: EMBL:U65431; NID:G2281290; PID:G2281291; PIDN:AAB64082.1  
C;Genetics:  
A;Gene: COLA2  
A;Cross-references: FlyBase:FBgn0016075  
C;Superfamily: collagen alpha 1(IV) chain

Query Match 4.5%; Score 11; DB 2; Length 1761;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCLEEFRA 184  
|||  
Db 1668 SPGSCLEEFRA 1678

RESULT 19  
CGH2B  
collagen alpha 2(IV) chain precursor - human  
N;Alternate names: procollagen alpha 2(IV) chain  
C;Species: Homo sapiens (man)  
C;Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C;Accession: A32024; S00007; S02624; S00246; S17678; S16511; B32117; S16877; S00165; S39  
R;Hostikka, S.L.; Tryggvason, K.  
J. Biol. Chem. 263, 19488-19499, 1988  
A;Title: The complete primary structure of the alpha2 chain of human type IV collagen an  
A;Reference number: A32024; MUID:89066769; PMID:3:98637  
A;Accession: A32024  
A;Molecule type: mRNA  
A;Residues: 1-1712 <HOS1>  
A;Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:G29550; PIDN:CAA29098.1; PI  
R;Hostikka, S.L.; Kurkinen, M.; Tryggvason, K.  
FEBS Lett. 216, 281-286, 1987

A;Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA re  
ated region.  
A;Reference number: S00007; MUID:87219158; PMID:3582677  
A;Accession: S00007  
A;Molecule type: mRNA  
A;Residues: 1254-1398, 'V', 1400-1712 <HOS2>  
A;Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:G29550; PIDN:CAA29098.1; PI  
A;Note: 1399-Ile was also found  
R;Hostikka, S.L.; Tryggvason, K.  
FEBS Lett. 224, 297-305, 1987  
A;Title: Extensive structural differences between genes for the alpha(1) and alpha(2) ch  
A;Reference number: S02624; MUID:88083553; PMID:2826228  
A;Accession: S02624  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1347-1350; 1377-1383; 1426-1432; 1465-1471; 1529-1535; 1625-1630 <HOS3>  
A;Note: complete nucleotide sequence not shown  
R;Brazel, D.; Pollner, R.; Oberbaumer, I.; Kuehn, K.  
Eur. J. Biochem. 172, 35-42, 1988  
A;Title: Human basement membrane collagen (type IV): the amino acid sequence of the alph  
A;Reference number: S00246; MUID:88151998; PMID:3345760  
A;Accession: S00246  
A;Molecule type: mRNA  
A;Residues: 1-682, 'G', 684-1043 <BRA>  
A;Cross-references: EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID:G30076  
R;Oberbaumer, I.  
submitted to the EMBL Data Library, June 1987  
A;Reference number: S17678  
A;Accession: S17678  
A;Molecule type: mRNA  
A;Residues: 1-470, 'P', 472-582, 'G', 684-1043 <OBE>  
A;Cross-references: EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID:G30076  
R;Poeschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c  
A;Reference number: S02738; MUID:89030632; PMID:2846280  
A;Accession: S16911  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-33 <POE>  
A;Cross-references: EMBL:X12784; GB:M36963; NID:G30072; PIDN:CAA31275.1; PID:G30073  
R;Solinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A;Reference number: A92850; MUID:89034231; PMID:3182844  
A;Accession: B32117  
A;Molecule type: DNA  
A;Residues: 1-33 <SO11>  
A;Cross-references: EMBL:J04217; EMBL:J05039; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R;Solinen, R.; Huotari, M.; Garguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll  
A;Reference number: S16876; MUID:89340433; PMID:2701944  
A;Accession: S16877  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-33 <SO12>  
A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233; EMBL:J05039  
A;Note: this sequence was submitted to the EMBL Data Library, October 1988  
R;Siebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, H.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 168, 569-575, 1987  
A;Title: Construction of a model for the aggregation and cross-linking region (7S domain  
is region.  
A;Reference number: S00165; MUID:88029476; PMID:3117548  
A;Accession: S00165  
A;Molecule type: protein  
A;Residues: 37-247 <SIE1>  
A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-Gly  
R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen  
A;Reference number: S39614; MUID:94038963; PMID:8223488  
A;Accession: S39615

A:Molecule type: protein  
A:Residues: 407-570 <EBL>  
R:MacWhright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (B)  
A:Reference number: S16910; MUID:84053346; PMID:6416291  
A:Accession: S16912  
A:Molecule type: protein  
A:Residues: 490-492, 'X', 494-496; 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>  
R:Experimental source: placenta  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A:Title: Pepsin fragments of human placental basement-membrane collagens showing intertr  
A:Reference number: S16908; MUID:82005835; PMID:6792033  
A:Accession: B58517  
A:Molecule type: protein  
A:Residues: 490-492, 'X', 494-501, 'P', 503-507; 952-957, 'X', 959-966, 'X', 968-984-986, 'X', 988-  
81-1185 <GLA>  
R:Killen, P.D.; Francomano, C.A.; Yamada, Y.; Modi, W.S.; O'Brien, S.J.  
Hum. Genet. 77, 318-324, 1987  
A:Title: Partial structure of the human alpha-2(IV) collagen chain and chromosomal local  
A:Reference number: S01450; MUID:88085168; PMID:3692475  
A:Accession: S01450  
A:Molecule type: mRNA  
A:Residues: 1040, 'L', 1042-1398, 'V', 1400-1418, 'M', 1420-1635, 'V', 1637-1712 <KIL>  
A:Cross-references: EMBL:M24766; NID:G537328; PIDN:AAA52043.1; PID:G537329  
R:Siebold, B.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 176, 617-624, 1988  
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxy-term  
A:Reference number: S02550; MUID:89005112; PMID:2844531  
A:Accession: S02550  
A:Molecule type: protein  
A:Residues: 1480-1535; 1545-1614; 1617-1662, 'H', 1664-1700, 'G', 1705-1708; 1710-1712 <SIE2>  
A:Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17  
R:Myers, J.C.; Howard, P.S.; Jelen, A.M.; Bion, A.S.; Macarak, E.J.  
J. Biol. Chem. 262, 9231-9238, 1987  
A:Title: Duplication of type IV collagen COOH-terminal repeats and species-specific exp  
A:Reference number: A27114; MUID:87250571; PMID:2439508  
A:Accession: B27114  
A:Molecule type: mRNA  
A:Residues: 1486-1574, 'I', 1576-1712 <MYE>  
A:Cross-references: EMBL:J02760; NID:G180425; PIDN:AAA58422.1; PID:G180426  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL4A2  
A:Cross-references: GDB:119792; OMIM:120090  
A:Map position: 13q34-13q34  
A:Introns: 15/2; 33/3; 134/71; 1380/1; 1429/1; 1458/1; 1532/1; 1527/3 #status incomplete  
A:Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with  
C:Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:COHU4B)  
domains (with disulfide and desmosine cross-links), dimeric associations among trimer ca  
rupted helical domain (with disulfide and desmosine cross-links).  
C:Function:  
A:Description: structural component of basement membrane  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glycopro  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1712/Product: collagen alpha 2(IV) chain #status predicted <MAT>  
F:29-57/Domain: amino-terminal nonhelical, NH1 <NH1>  
F:58-1485/Region: interrupted helical  
F:362-364/Region: cell attachment (R-G-D) motif  
F:784-786/Region: cell attachment (R-G-D) motif  
F:868-891/Region: cell attachment (R-G-D) motif  
F:889-891/Region: cell attachment (R-G-D) motif  
F:970-972/Region: cell attachment (R-G-D) motif  
F:1069-1071/Region: cell attachment (R-G-D) motif  
F:1228-1230/Region: cell attachment (R-G-D) motif  
F:1452-1454/Region: cell attachment (R-G-D) motif  
F:1486-1712/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F:1495-1593/Domain: collagen IV carboxyl-terminal repeat <CTR1>  
F:1603-1708/Domain: collagen IV carboxyl-terminal repeat <CTR2>  
F:42, 47, 51, 53, 137, 483, 485/Disulfide bonds: interchain #status predicted

P:57,87,90,102,165,168,225,239,242/Binding site: carbohydrate (Lys) (covalent) #status I  
F:57/Modified site: 5-hydroxylysine (Lys) #status atypical  
F:63,75,96,114,120,123,132,150,159,186,189,198,201,213,216,219,496,499,955,964,1103,1111  
F:87,90,102,165,168,225,239,242/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:138/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:209/Modified site: 4-hydroxyproline (Pro) #status atypical  
F:661-681/Disulfide bonds: #status predicted  
F:1275/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:1549-1555,1658-1665/Disulfide bonds: (or 1504-1593, 1537-1590) #status experimental  
F:1612-1705,1646-1708/Disulfide bonds: (or 1612-1708, 1646-1705) #status experimental

Query Match 4.1%; Score 10; DB 1; Length 1712;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQ 137  
|||||  
Db 1596 PAIAIAVHSQ 1605

RESULT 20  
A:54121  
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)  
N:Alternate names: collagen alpha 2(IV) chain homolog  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 13-Aug-1999  
C:Accession: A54121; S4317  
R:Exposito, J.F.; Suzuki, H.; Georjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.  
J. Biol. Chem. 269, 13167-13171, 1994  
A:Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(1  
A:Reference number: A54121; MUID:94230414; PMID:8175744  
A:Accession: A54121  
A:Molecule type: mRNA  
A:Residues: 1-1747 <EXP>  
A:Cross-references: EMBL:X76730; NID:G483606; PIDN:CAA54146.1; PID:G483607  
C:Genetics:  
A:Gene: COLP4alpha  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 4.1%; Score 10; DB 2; Length 1747;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 PYISRCTVCE 126  
|||||  
Db 1622 PYISRCTVCE 1631

RESULT 21  
I48303  
collagen alpha 4(IV) chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 13-Aug-1999  
C:Accession: I48303; S47279  
R:Miner, J.H.; Sanes, J.R.  
J. Cell Biol. 127, 879-891, 1994  
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequ  
A:Reference number: A54979; MUID:95050957; PMID:7962065  
A:Accession: I48303  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-312 <RES>  
A:Cross-references: EMBL:Z35167; NID:G535199; PIDN:CAA84530.1; PID:G535200  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 3.7%; Score 9; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 0.47;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ISRCQVCMK 241  
|||||  
Db 302 ISRCQVCMK 310

## RESULT 22

A61228  
collagen alpha 2(IV) chain precursor - rabbit (fragments)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 17-Mar-1999  
C/Accession: A61228  
R;Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.  
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991  
A>Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium  
A/Reference number: A61228; MUID:92010685; PMID:1717398  
A/Accession: A61228  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-775 <YAM>  
C/Superfamily: collagen alpha 1(IV) chain  
C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 3.7% Score 9; DB 2; Length 775;  
Best Local Similarity 100.0%; Pred No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ISRCQVCMK 241

|||||

Db 765 ISRCQVCMK 773

## RESULT 23

A33526  
collagen alpha 2(IV) chain precursor - mouse  
C:Species: Mus musculus (house mouse)  
C/Date: 04-Dec-1992 #sequence\_revision 13-Aug-1999  
C/Accession: A33526; A24432; B25066; A24364; S19081; B29301; A24628; B32  
E;Saus, J.; Quinones, S.; Mackrell, A.; Blumberg, B.; Muthukumar, G.; Pihlajaniemi, T.  
J. Biol. Chem. 264, 6318-6324, 1989  
A>Title: The complete primary structure of mouse alpha-2(IV) collagen. Alignment with mo  
A/Reference number: A33526; MUID:89197933; PMID:2703491  
A/Accession: A33526  
A/Molecule type: mRNA  
A/Residues: 1-1707 <SAU>  
A/Cross-references: EMBL:J04695; NID:9556298; PIDN:AAA50293.1; PID:9556299  
R;Kurkinen, M.; Bernard, M.P.; Barlow, D.P.; Chow, L.T.  
Nature 317, 177-179, 1985  
A>Title: Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha-2(IV)  
A/Reference number: A93367; MUID:85296379; PMID:3839908  
A/Accession: A24432  
A/Molecule type: mRNA  
A/Residues: 967-1096 'G', 1098-1109 <KUL>  
A/Cross-references: EMBL:X02896; NID:950263; PIDN:CAA26655.1; PID:9899326  
A/Note: The authors translated the codon AAC for residue 964 as Lys  
A/Accession: D24432  
A/Molecule type: DNA  
A/Residues: 964-1096 'G', 1098-1109 <KUL>  
A/Cross-references: EMBL:X02899  
R;Schwarz, U.; Schuppan, D.; Oberbauer, I.; Glanville, R.W.; Deutzmann, R.; Timpl, R.;  
Eur. J. Biochem. 157, 49-56, 1986  
A>Title: Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-  
A/Reference number: A25066; MUID:86220192; PMID:3011432  
A/Accession: A25066  
A/Molecule type: mRNA  
A/Residues: 970-1480 <SCI>  
A/Cross-references: EMBL:X04647  
A/Accession: B25066  
A/Molecule type: protein  
A/Residues: 979-1058;1065-1101;1105-1222;1226-1310;1326-1335;1351-1480 <SC2>  
R;Vogeli, G.; Horn, E.; Carter, J.; Kaytes, P.S.  
FEBS Lett. 206, 29-32, 1986  
A>Title: Proposed alignment of helical interruptions in the two subunits of the basement  
A/Reference number: A24364; MUID:87005245; PMID:3758345  
A/Accession: A24364  
A/Molecule type: mRNA  
A/Residues: 1041-1050, 'R', 1052-1170, 'S', 1172-1178, 'R', 1180-1240, 'E', 1242-1327, 'A', 1329-1

A/Cross-references: EMBL:X04410; NID:950240; PIDN:CAA27998.1; PID:9929678  
R;Kaytes, P.S.; Theriault, N.Y.; Vogeli, G.  
Gene 54, 141-146, 1987  
A>Title: Homologies between the non-collagenous C-terminal (NC1) globular domains of the  
A/Reference number: S19080; MUID:87277427; PMID:3609751  
A/Accession: S19081  
A/Molecule type: mRNA  
A/Residues: 1466-1622 'H', 1624-1707 <KAI>  
A/Cross-references: GB:X04410; NID:950240; PIDN:CAA27998.1; PID:9929678  
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj  
J. Biol. Chem. 262, 8496-8499, 1987  
A>Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
A/Reference number: A94680; MUID:87250460; PMID:3597383  
A/Accession: B29301  
A/Molecule type: mRNA  
A/Residues: 1481-1707 <KUR>  
A/Cross-references: EMBL:M15833; NID:912284; PIDN:AAA37341.1; PID:9387116  
R;Schwarz-Magdolen, U.; Oberbauer, I.; Kuehn, K.  
FEBS Lett. 208, 203-207, 1986  
A>Title: cDNA and protein sequence of the NC1 domain of the alpha-2-chain of collagen IV  
A/Reference number: A24628; MUID:87054581; PMID:3780963  
A/Accession: A24628  
A/Molecule type: mRNA  
A/Residues: 1480-1572, 'L', 1574-1622, 'H', 1624-1707 <SCH>  
A/Cross-references: EMBL:X04647  
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
J. Biol. Chem. 263, 19274-19277, 1988  
A>Title: Head-to-head arrangement of murine type IV collagen genes.  
A/Reference number: A92702; MUID:89066738; PMID:3198626  
A/Accession: B32003  
A/Molecule type: DNA  
A/Residues: 1-33 <KA2>  
A/Cross-references: EMBL:J04448; NID:912666; PIDN:AAA37438.1; PID:9126667  
R;Burbello, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A>Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
A/Reference number: A94220; MUID:89071759; PMID:3200851  
A/Accession: B31786  
A/Molecule type: DNA  
A/Residues: 1-60 <BUR>  
A/Cross-references: EMBL:M23333  
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 139, 401-410, 1984  
A>Title: Subunit structure and assembly of the globular domain of basement-membrane coll  
A/Reference number: S17801; MUID:84132058; PMID:6698021  
A/Accession: S19086  
A/Molecule type: protein  
A/Residues: 1475-1481, 'X', 1483-1487 <WEB>  
C/Genetics:  
A/Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3  
C/Superfamily: collagen alpha 1(IV) chain  
C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F:29-171/Domain: 7S #status predicted <7SD>  
F:58-1480/Domain: collagenous #status predicted <COL>  
F:141-143/Region: cell attachment (R-G-D) motif  
F:360-362/Region: cell attachment (R-G-D) motif  
F:779-781/Region: cell attachment (R-G-D) motif  
F:884-886/Region: cell attachment (R-G-D) motif  
F:965-967/Region: cell attachment (R-G-D) motif  
F:1223-1225/Region: cell attachment (R-G-D) motif  
F:1447-1449/Region: cell attachment (R-G-D) motif  
F:1481-1707/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F:1481-1589/Domain: repeat NC1 #status predicted <NC1>  
F:1590-1707/Domain: repeat NC1 #status predicted <NC1>  
F:42, 47, 51, 53, 481, 483/Disulfide bonds: interchain #status predicted  
F:138, 1270/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:656-676, 1544-1550, 1653-1660/disulfide bonds: #status predicted  
F:985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Bind  
F:985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Mod  
F:1009, 1012, 1018, 1021, 1024, 1037, 1040, 1043, 1046, 1052, 1058, 1070, 1098, 1110, 1128, 1140, 1149, 1  
77, 1383, 1386, 1401, 1408, 1420, 1423, 1429, 1444, 1465, 1468, 1471, 1477/Modified site: hydroxypr

C:Keywords: iron; manganese; metalloprotein; oxidoreductase  
 F:104,152,236,240/Binding site: iron/manganese (His, His, Asp, His) #status predicted  
 3.3%; Score 8; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred.No. 4.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 114 ALEPYISR 121  
 |||||  
 DB 91 ALEPYISR 98  
 RESULT 26  
 CGHU6B  
 collagen alpha 6(IV) chain precursor - human  
 N;Alternate names: procollagen alpha 6(IV) chain  
 C;Species: Homo sapiens (man)  
 C;Date: 07-Jul-1995 #sequence revision 03-Oct-1995 #text\_change 16-Jun-2000  
 C;Accession: A54122; A53404; B57079  
 R;Zhou, J.; Ding, M.; Zhao, Z.; Reeders, S.T.  
 J. Biol. Chem. 269, 13193-13199, 1994  
 A;Title: Complete primary structure of the sixth chain of human basement membrane collagen  
 A;Reference number: A54122; MUID:94230418; PMID:8175748  
 A;Accession: A54122  
 A;Molecule type: mRNA  
 A;Residues: 1-1691 <ZHO>  
 A;Cross-references: GB:U04845; NID:g496977; PIDN:AAAL9569.1; PID:g496978  
 R;Oohashi, T.; Sugamoto, M.; Mattei, M.G.; Ninomiya, Y.  
 J. Biol. Chem. 269, 7520-7526, 1994  
 A;Title: Identification of a new collagen IV chain, alpha6(IV), by cDNA isolation and as  
 A;Reference number: A53404; MUID:94171779; PMID:8125972

A:Accession: A53404  
A:Molecule type: mRNA  
A:Residues: 'MHQ', 6-169, 'M', 171-916, 'S', 918-1301, 1314-1355, 'A', 1357-1691 <OOH>  
A:Cross-references: DDBJ: D213137; NID: G466337; PIDN: BAA04809.1; PID: G466538  
R: Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurilla, P.; de Paeppe, A.; Tryggvason, R. 261, 1167-1169, 1993  
A:Title: Deletion of the paired alphas(IV) and alphas(IV) collagen genes in inherited sr  
A:Reference number: A57079; UID: 93361972; PMID: 8356449

A;Accession: B37079  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-546,'G','ZH2'  
A;Cross-references: GB:L22763  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit ( )  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL4A6  
A;Cross-references: GDB:222775; OMIM:303631  
A;Map position: Xq22-Xq22  
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with  
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5(  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a  
er associations in the interrupted helical domain (with disulfide and desmosine cross-li  
C;Function:  
A;Description: minor structural component of extracellular basement membrane  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; Glyc  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-1691/Product: collagen alpha 6(IV) chain #status predicted <MAT>  
F;23-46/Domain: amino-terminal nonhelical, NC2 <NC2>  
F;47-1463/Region: interrupted helical  
F;F;1464-1691/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F;1473-1571/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1581-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F;31,36,40,42,126,482,484,657/Disulfide bonds: interchain #status predicted  
F;127/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1482-1568,1515-1571/Disulfide bonds: (or 1482-1571, 1515-1568) #status predicted  
F;1527-1533,1636-1643/Disulfide bonds: #status predicted  
F;1590-1684,1624-1687/Disulfide bonds: (or 1590-1687, 1624-1684) #status predicted

```

F:1464-1691/Domain: carboxyl-terminal nonhelical, NCI <NC1>
F:1473-1571/Domain: collagen IV carboxyl-terminal repeat <CRL>
F:1581-1687/Domain: collagen IV carboxyl-terminal repeat <CR2>
P:31-36,40,42,126,482,484,657/Disulfide bonds: interChain #status predicted
P:147-binding site: carboxylate (Asn) (covalent) #status predicted
P:1482-1568,1515-1571/Disulfide bonds: (or 1482-1571, 1515-1568) #status predicted
P:1527-1533,1636-1643/Disulfide bonds: #status predicted
P:1590-1684,1624-1687/Disulfide bonds: (or 1590-1687, 1624-1684) #status predicted

Query Match 3.3%; Score 8; DB 1; Length 1691;

```

Best Local Similarity 100.0%; Pred. No. 20; Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 234 SRCQVCMK 241  
Db 1682 SRCQVCMK 1689

RESULT 27  
A31893 collagen alpha 1(IV) chain precursor - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C>Date: 21-May-1990 #sequence\_revision 21-May-1990 #text\_change 21-Jul-2000  
C:Accession: A31893; A26692; A19442; S00020  
R:Blumberg, B.; Mackrell, A.J.; Fesseler, J.H.  
J. Biol. Chem. 263, 18328-18337, 1988  
A:Title: *Drosophila* basement membrane procollagen alpha-1(IV). II. Complete cDNA sequence  
A:Reference number: A31893; MUID:89054012; PMID:3142875  
A:Accession: A31893  
A:Molecule type: mRNA  
A:Residues: 1-1775 <BLU>  
A:Cross-references: EMBL:W23704; NID:G157029; PIDN:AAA28404.1; PID:G157030  
R:Blumberg, B.; Mackrell, A.J.; Olson, F.F.; Kurkinen, M.; Monson, J.M.; Natzle, J.E.; P.  
J. Biol. Chem. 262, 5947-5950, 1987  
A:Title: Basement membrane procollagen IV and its specialized carboxyl domain are conserved  
A:Reference number: A26692; MUID:87194801; PMID:3106346  
A:Accession: A26692  
A:Molecule type: mRNA  
A:Residues: 1065-1775 <BLU2>  
A:Cross-references: EMBL:J02727  
R:Monson, J.M.; Natzle, J.; Friedman, J.; McCarthy, B.J.  
Proc. Natl. Acad. Sci. U.S.A. 79, 1761-1765, 1982  
A:Title: Expression and novel structure of a collagen gene in *Drosophila*.  
A:Reference number: A19442; MUID:82197577; PMID:6210912  
A:Accession: A19442  
A:Molecule type: DNA  
A:Residues: 762-947, 'S', 949-996, 'T', 998-1230 <MON>  
A:Cross-references: GB:J01074; EMBL:V00200; NID:G7736; PIDN:CAA23486.2; PID:G5777391  
R:Cecchini, J.P.; Knibbs, B.; Mirre, C.; le Parco, Y.  
Eur. J. Biochem. 165, 587-593, 1987  
A:Title: Evidence for a type-IV-related collagen in *Drosophila melanogaster*. Evolutionary  
A:Reference number: S00020; MUID:87246644; PMID:3109906  
A:Accession: S00020  
A:Molecule type: DNA  
A:Residues: 1355-1356, 'K', 1358-1359, 'K', 1361-1372, 'I', 1374-1495, 'R', 1497-1506, 'RA', 1509,  
A:Cross-references: EMBL:M28334  
C:Genetics:  
A:Gene: FlyBase:CG25C  
A:Cross-references: FlyBase:FBgn0000299  
A:Introns: 7/2; 23/3; 339/3; 505/2; 999/1; 1312/1; 1699/3  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-1775/Product: collagen alpha 1(IV) chain #status predicted <NAT>  
F:65-67/Region: cell attachment (R-G-D) motif  
F:130-132/Region: cell attachment (R-G-D) motif  
F:238-240/Region: cell attachment (R-G-D) motif  
F:297-299/Region: cell attachment (R-G-D) motif  
F:892-894/Region: cell attachment (R-G-D) motif  
F:1075-1077/Region: cell attachment (R-G-D) motif  
F:1173-1175/Region: cell attachment (R-G-D) motif  
F:1225-1227/Region: cell attachment (R-G-D) motif  
F:1545-1775/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>  
F:1545-1655/Domain: repeat NCI #status predicted <NC11>  
F:1656-1775/Domain: repeat NCI #status predicted <NC12>  
F:72/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:570, 573/Disulfide bonds: interchain #status predicted  
F:1611-1617, 1720-1727/Disulfide bonds: #status predicted

Query Match 3.3%; Score 8; DB 2; Length 1775;  
Best Local Similarity 100.0%; Pred. No. 21; Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 234 SRCQVCMK 241  
Db 1765 SRCQVCMK 1772

RESULT 28  
S50999 superoxide dismutase (EC 1.15.1.1) (Fe) - *Azotobacter vinelandii* (fragment)  
C:Species: *Azotobacter vinelandii*  
C>Date: 15-Jul-1995 #sequence\_revision 14-Nov-1997 #text\_change 05-Mar-1999  
C:Accession: S50999  
R:Pagan, S.; Colnaghi, R.; Palagi, A.; Negri, A.  
FEBS Lett. 357, 79-82, 1995  
A:Title: Purification and characterization of an iron superoxide dismutase from the nitr  
A:Reference number: S50999; MUID:95094938; PMID:8001685  
A:Accession: S50999  
A:Molecule type: protein  
A:Residues: 1-49 <PAG>  
A:Experimental source: strain UW136  
C:Function:  
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
C:Superfamily: superoxide dismutase (Mn)  
C:Keywords: homodimer; metalloprotein; oxidoreductase

Query Match 2.9%; Score 7; DB 2; Length 49;  
Best Local Similarity 100.0%; Pred. No. 11; Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 114 ALBPYIS 120  
Db 13 ALBPYIS 19

RESULT 29  
C72639 hypothetical protein AP0547 - *Aeropyrum pernix* (strain K1)  
C:Species: *Aeropyrum pernix*  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C:Accession: C72639  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: C72639  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-122 <KAW>  
A:Cross-references: DDBJ:AP0000060; NID:G5104188; PIDN:BAA79515.1; PID:d1043301; PID:G510  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: AP0547  
C:Superfamily: *Aeropyrum pernix* hypothetical protein AP0547

Query Match 2.9%; Score 7; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 23; Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 60 DLGTILGS 66  
Db 52 DLGTILGS 58

RESULT 30  
B42526 B3R protein - vaccinia virus (strain Copenhagen)  
C:Species: vaccinia virus  
A:Note: host *Homo sapiens* (man)  
C>Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 08-Apr-1994  
C:Accession: B42526  
R:Johnson, G.P.  
submitted to GenBank, June 1990  
A:Reference number: A33172



Query Match 2.9%; Score 7; DB 2; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 228 ELEKIIS 234

DB 61 ELEKIIS 67

# RESULT 36

A87671  
 cytochrome c oxidase assembly protein, probable [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: A87671  
 R:Nierman, W.C.; Fejdyblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Leub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; MUID:21173698; PMID:11259647  
 A:Accession: A87671  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-200 <STO>  
 A:Cross-references: GB:AE005673; NID:g13425113; PIDN:AAK25365.1; GSPDB:GN00148  
 C:Genetics:  
 A:Gene: CC3403  
 C:Superfamily: cytochrome-c oxidase assembly protein COX11

Query Match 2.9%; Score 7; DB 2; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115

DB 101 PITGRAL 107

# RESULT 37

S51097  
 superoxide dismutase (EC 1.15.1.1) (Fe/Mn) - Methanobacterium thermoautotrophicum (strai  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Apr-2000  
 C:Accession: S51097  
 R:Maile, L.; Fischer, K.; Jenal, U.; Leisinger, T.  
 submitted to the EMBL Data Library, July 1993  
 A:Description: Molecular characterization of a superoxide dismutase gene from Methanobac  
 A:Reference number: S51097  
 A:Accession: S51097  
 A:Molecule type: DNA  
 A:Residues: 1-202 <MEI>  
 A:Cross-references: EMBL:X74264; NID:g620124; PIDN:CAA52323.1; PID:g620125  
 A:Experimental source: strain Marburg, DSM 2133  
 C:Genetics:  
 A:Gene: sod  
 C:Function:  
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen

C:Superfamily: superoxide dismutase (Mn)  
 C:Keywords: iron; manganese; metalloprotein; oxidoreductase  
 F:30,78,164,168/Binding site: iron/manganese (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120

DB 17 ALEPYIS 23

# RESULT 38

JC4396  
 superoxide dismutase (EC 1.15.1.1) (Fe/Mn) [validated] - Propionibacterium freudenreichi  
 C:Species: Propionibacterium freudenreichii subsp. shermanii  
 C>Date: 20-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 20-Apr-2000  
 C:Accession: JC4396; S41106  
 R:Gabbianelli, R.; Battistoni, A.; Polizio, F.; Carri, M.T.; De Martino, A.; Meier, B.;  
 Biochem. Biophys. Res. Commun. 216, 841-847, 1995  
 A:Title: Metal uptake of recombinant cambialistic superoxide dismutase from Propionibac  
 A:Reference number: JC4396; MUID:96074560; PMID:7488202  
 A:Accession: JC4396  
 A:Molecule type: DNA  
 A:Residues: 1-202 <GAB>  
 A:Cross-references: EMBL:X91650  
 A:Experimental source: PZ3  
 R:Meier, B.; Sehm, A.F.; Schinina, M.E.; Barra, D.  
 Eur. J. Biochem. 219, 463-468, 1994  
 A:Title: In vivo incorporation of copper into the iron-exchangeable and manganese-exch  
 A:Reference number: S41106; MUID:94139724; PMID:8307013  
 A:Accession: S41106  
 A:Molecule type: protein  
 A:Residues: 2-202 <MEI>  
 A:Experimental source: strain PZ3  
 C:Genetics:  
 A:Gene: sod  
 C:Complex: homotetramer  
 C:Function:  
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyge  
 A:Note: can use iron or manganese as cofactor  
 C:Superfamily: superoxide dismutase (Mn)  
 C:Keywords: homotetramer; iron; manganese; metalloprotein; oxidoreductase  
 F:2-202/Product: superoxide dismutase #status experimental <MAT>  
 F:28,76,162,166/Binding site: iron/manganese (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120

DB 15 ALEPYIS 21

# RESULT 39

E90174  
 superoxide dismutase [Fe] (sod) [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
 C:Accession: E90174  
 R:Singh, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Cha  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 arett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: E90174  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-211 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13813460; PIDN:AAK40652.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: sod  
 C:Superfamily: superoxide dismutase (Mn)

Query Match 2.9%; Score 7; DB 2; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120

DB 21 ALEPYIS 27



**RESULT 42**

E81247

cell division ATP-binding protein FtsE NMB0007 [imported] - Neisseria meningitidis (strain E81247)

C:Species: *Neisseria meningitidis*

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: E81247

R:Terteljin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Wickett, B.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Xie, H.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlata, V.; Masiognani, V.; Pizzo, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.; Venkatesan, A.

A>Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A:Reference number: AB1000; MUID:20175755; PMID:10710307

A:Accession: E81247

C:Genetic8;  
A:Gene: SGD:SD2; MIPS:YHR008C  
A:Cross-references: SGD:S0001050; MIPS:YHR008C  
A:Map position: BR  
A:Genome: nuclear  
C:Complex: homotetramer  
C:Function:  
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
C:Superfamily: superoxide dismutase (Mn)  
C:Keywords: homotetramer; manganese; metalloprotein; mitochondrial matrix; mitochondrion  
F:1-26/Domain: transit peptide (mitochondrion) #status predicted <RNP>  
F:127-229/Product: superoxide dismutase (Mn) #status experimental <MAT>  
F:252-194/Binding site: manganese (His, Asp, His) #status predicted  
P:527-194/Binding site: manganese (His, Asp, His) #status predicted

RESULT 44  
H84035

hypothetical protein BH3088 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: H84035  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512592; PMID:11059132  
 A:Accession: H84035  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-237 <STO>  
 A:Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA806807.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 C:Superfamily: Bacillus subtilis hypothetical protein yoaT

Query Match 2.9%; Score 7; DB 2; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VPLYSGF 45  
 |||||  
 Db 100 VPLYSGF 106

RESULT 45  
 AB0520  
 conserved hypothetical protein STY0161 [imported] - Salmonella enterica subsp. enterica  
 C:Species: Salmonella enterica subsp. enterica serovar typhi  
 A:Note: This species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AB0520  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 T.; T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Fairy, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AB0520  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-247 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD01298.1; PID:G16501426; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0161  
 C:Superfamily: Escherichia coli hypothetical protein b0102

Query Match 2.9%; Score 7; DB 2; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 WLASLNP 212  
 |||||  
 Db 157 WLASLNP 163

RESULT 46  
 JG0179  
 superoxide dismutase (EC 1.15.1.1) (Fe) - rice  
 C:Species: Oryza sativa (rice)  
 C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: JG0179  
 R:Kaminaka, H.; Morita, S.; Tokumoto, M.; Yokoyama, H.; Masumura, T.; Tanaka, K.  
 Biosci. Biotechnol. Biochem. 63, 302-308, 1999  
 A:Title: Molecular cloning and characterization of a cDNA for an iron-superoxide dismuta  
 A:Reference number: JG0179; MUID:99208990; PMID:10192910  
 A:Accession: JG0179  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-255 <KAM>

A:Cross-references: DDBJ:AB014056; NID:G4164148; PIDN:BAA37131.1; PID:G4164149  
 C:Superfamily: superoxide dismutase (Mn)  
 C:Keywords: iron; metalloprotein; oxidoreductase  
 F:67,119,203; Binding site: iron (His, His, Asp, His) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALPEYIS 120  
 |||||  
 Db 54 ALPEYIS 60

RESULT 47  
 H82575  
 3-deoxy-manno-octulosonate cytidylyltransferase XF2299 [imported] - Xylella fastidiosa (C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: H82575  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20355717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82575  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-257 <SIM>  
 A:Cross-references: GB:AE004041; GB:AE003849; NID:G9107453; PIDN:AAF85098.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
 as-Neto, E.; Docena, C.; El-Dosry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 Submitted to GenBank, June 2000  
 A:Authors: Ferreira, J.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.C.R.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2299  
 C:Superfamily: 3-deoxy-manno-octulosonate cytidylyltransferase

Query Match 2.9%; Score 7; DB 2; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRTTM 74  
 |||||  
 Db 199 LQRTTM 205

RESULT 48  
 WMS28  
 complement factor D (EC 3.4.21.46) precursor - mouse  
 N:Alternate names: adipocyte 28k proteinase; adipsin; C3 convertase activator; compleme  
 C:Species: Mus musculus (house mouse)  
 C>Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 19-May-2000  
 C:Accession: C25952; A00937; A26105  
 R:Phillips, M.; Djan, P.; Green, H.  
 J. Biol. Chem. 261, 10821-10827, 1986  
 A:Title: The nucleotide sequence of three genes participating in the adipose differentia  
 A:Reference number: A92553; MUID:86278164; PMID:3015943  
 A:Accession: C25952  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Residues: 1-259 <PHI>  
 A:Cross-references: GB:ML3386; NID:G192033; PIDN:AAA37262.1; PID:G387105

R;Cook, K.S.; Groves, D.L.; Min, H.Y.; Spiegelman, B.M.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6480-6484, 1985  
A;Title: A developmentally regulated mRNA from 3T3 adipocytes encodes a novel serine protease  
A;Reference number: A00937; MUID:86016726; PMID:3901003  
A;Accession: A00937  
A;Molecule type: mRNA  
A;Residues: 1-259 <COO>  
A;Cross-references: GB:M11768; NID:G202166; PIDN:AAA40486.1; PID:G202167  
A;Experimental source: strain Swiss White  
A;Note: Only one Ala is present in place of Ala-19 and Ala-20 in another equally abundant  
R;Min, H.Y.; Spiegelman, B.M.  
Nucleic Acids Res. 14, 8878-8892, 1986  
A;Title: Adipin, the adipocyte serine protease: gene structure and control of expression  
A;Reference number: A26105; MUID:87066764; PMID:3024123  
A;Accession: A26105  
A;Molecule type: mRNA  
A;Residues: 1-259 <MIN>  
A;Cross-references: GB:M11768; NID:G493883; PIDN:CAA28378.1; PID:G581866  
C;Comment: Human complement factor D is synthesized primarily in cells of the macrophage  
C;Comment: activation. However, expression of the murine homolog is specific to adipose tissue  
C;Genetics:  
A;Introns: 19/1; 71/2; 120/3; 206/3  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: adipose tissue; alternative splicing; hydrolase; serine proteinase  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-25/Domain: propeptide #status predicted <PRO>  
F;26-259/Product: adipin #status predicted <MAT>  
F;26-249/Domain: trypsin homology <TRY>  
F;51-67/149-215,180-196,205-230/Disulfide bonds: #status predicted  
F;66,115,209/Active site: His, Asp, Ser #status predicted

Query Match 2.9%; Score 7; DB 1; Length 259;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSP 12  
|||||  
DB 206 RGDGSP 212

RESULT 49  
A34476  
collagen alpha 2(IV) chain - Caenorhabditis elegans (fragment)  
C;Species: Caenorhabditis elegans  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: A34476  
R;Guo, X.; Kramer, J.M.  
J. Biol. Chem. 264, 17574-17582, 1989  
A;Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1  
A;Reference number: A34476; MUID:90008929; PMID:2793871  
A;Accession: A34476  
A;Molecule type: DNA  
A;Residues: 1-261 <GUO>  
A;Cross-references: EMBL:J05066; NID:G156259; PIDN:AAA27989.1; PID:G156260  
C;Genetics:  
A;Gene: Clb1  
A;Map position: X  
A;Introns: 77/1; 231/3  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;  
F;1-30/Domain: collagenous (fragment) #status predicted <COL>  
F;31-261/Domain: carboxyl-terminal nonhelical, NCL1 #status predicted <NCL1>  
F;31-139/Domain: repeat NCL1 #status predicted <NCL1>  
F;140-261/Domain: repeat NCL1 #status predicted <NCL1>  
F;94-100,203-210/Disulfide bonds: #status predicted

Query Match 2.9%; Score 7; DB 2; Length 261;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLQRF 71  
|||||

DB 80 GSCLQRF 86

RESULT 50  
E69897  
hypothetical protein yoar - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C;Accession: E69897  
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:  
C;Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle:  
A;Jech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue:  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Postecell:  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon:  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron:  
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama:  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, J  
A;Authors: Yoshikawa, H.P.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:96044033; PMID:9384377  
A;Accession: E69897  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Residues: 1-264 <KUN>  
A;Molecule type: DNA  
A;Cross-references: GB:Z99114; GB:AL009126; NID:G2634230; PIDN:CAB13767.1; PID:G2634268  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: yoar  
C;Superfamily: Bacillus subtilis hypothetical protein yoar

Query Match 2.9%; Score 7; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45  
|||||  
DB 103 VPLYSGF 109

RESULT 51  
AD1561  
B. subtilis yoar protein homolog lin1029 [imported] - Listeria innocua (strain Clip1126:  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AD1561  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker:  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; M:  
ok, C.; Schluter, T.; Smeets, N.; Tierrez, A.; Vazquez-Boland, J.A.; Vose, H.; Wehland:  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: A31077; MUID:21537279; PMID:11679669  
A;Accession: AD1561  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-267 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC96260.1; PID:G16413488; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin1029  
C;Superfamily: Bacillus subtilis hypothetical protein yoar

Query Match 2.9%; Score 7; DB 2; Length 267;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45

Db 103 VPLYSGF 109  
|||||

## RESULT 52

AE1204  
B. subtilis YoaT protein homolog lmo1037 [imported] - Listeria monocytogenes (strain EGD)  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AE1204  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluster, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1204  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99115.1; PID:gl6410439; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1037  
A:Superfamily: Bacillus subtilis hypothetical protein yoaT

Query Match 2.9%; Score 7; DB 2; Length 267;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 VPLYSGF 45  
|||||  
Db 103 VPLYSGF 109

## RESULT 53

AB6241  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C:Accession: AB6241  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719; PMID:11130712  
A:Accession: AB6241  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-343 <STO>  
A:Cross-references: GB:AE005172; NID:g4874273; PIDN:AAD31338.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: Arabidopsis thaliana hypothetical protein F16G20.50

Query Match 2.9%; Score 7; DB 2; Length 343;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 34 CPEGTVP 40  
|||||  
Db 55 CPEGTVP 61

## RESULT 54

AB6241

## C83577

hypothetical protein PA0549 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83577  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: C83577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-354 <STO>  
A:Cross-references: GB:AE004491; GB:AE004091; NID:g9946412; PIDN:AAG03938.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0549

Query Match 2.9%; Score 7; DB 2; Length 354;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 174 SPGSCLE 180  
|||||  
Db 115 SPGSCLE 121

## RESULT 55

SS2699  
hypothetical protein YDR191w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YD9346.03  
C:Species: Saccharomyces cerevisiae  
C>Date: 19-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Oct-1999  
C:Accession: SS2699  
R:Oliver, K.; Harris, D.  
submitted to the EMBL Data Library, March 1995  
A:Reference number: SS2699  
A:Accession: SS2699  
A:Molecule type: DNA  
A:Residues: 1-370 <OLI>  
A:Cross-references: EMBL:Z48784; NID:g755782; PIDN:CAA88705.1; PID:g755785; MIPS:YDR191w  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:HST4  
A:Cross-references: SGD:S0002599; MIPS:YDR191w  
A:Map position: 4R

Query Match 2.9%; Score 7; DB 2; Length 370;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 218 KPFPSTV 224  
|||||  
Db 204 KPFPSTV 210

## RESULT 56

SS7689  
hypothetical protein YGR225w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G8541  
C:Species: Saccharomyces cerevisiae  
C>Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 19-Apr-2002  
C:Accession: SS7689; S64549; S63905  
R:van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.  
submitted to the EMBL Data Library, June 1995  
A:Description: Sequence analysis of the 43 KB CRM1-YLM9-PET54-SMT1-PHO81-YHB4-PFK1 regic  
A:Reference number: SS7689  
A:Accession: SS7689  
A:Molecule type: DNA  
A:Residues: 1-409 <VAN>  
A:Cross-references: EMBL:X87941; NID:G886908; PIDN:CAA61174.1; PID:G886918

A;Experimental source: strain S288C  
R;van der Aart, Q.J.M.; Steensma, H.Y.  
A;Reference number: S64541  
A;Accession: S64549  
A;Molecule type: DNA  
A;Residues: 1-409 <VAV>  
A;Cross-references: EMBL:Z73010; NID:G1323405; PIDN:CAA97253.1; PID:e243661; PID:G132340  
A;Experimental source: strain S288C  
R;van der Aart, Q.J.M.; Kline, K.; Steensma, H.Y.  
Yeast 12, 385-390, 1996  
A;Title: Sequence analysis of the 43 kb CRM1-YLM3-PFT54-DIE2-SM11-PHO81-YHB4-PFK1 region  
A;Reference number: S63896; MUID:96267763; PMID:6701610  
A;Accession: S63905  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-409 <VAF>  
A;Cross-references: EMBL:X87941; NID:G886908; PIDN:CAA61174.1; PID:G886918  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995  
C;Genetics:  
A;Gene: SGD:AMA1  
A;Cross-references: SGD:S0003457  
A;Map position: 7R  
A;Note: YGR225w

Query Match 2.9%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 ASRNDYS 94  
|||||

Db 112 ASRNDYS 118  
|||||

## RESULT 57

S41607  
atrolysin A (EC 3.4.24.1) - western diamondback rattlesnake (fragment)  
N;Alternate names: hemorrhagic toxin a  
C;Species: Crotalus atrox (western diamondback rattlesnake)  
C;Date: 29-Sep-1994 #sequence\_revision 13-Mar-1997 #text\_change 09-Jun-2000  
C;Accession: S41607  
R;Hite, L.A.; Jia, L.G.; Bjarnason, J.B.; Fox, J.M.  
Arch. Biochem. Biophys. 308, 182-191, 1994  
A;Title: cDNA sequences for four snake venom metalloproteinases: structure, classification  
A;Reference number: S41607; MUID:94145078; PMID:8311451  
A;Accession: S41607  
A;Status: preliminary; translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-419 <HIT>  
A;Cross-references: EMBL:U01234; NID:G402257; PID:G402258  
C;Superfamily: mouse meltrin alpha; disintegrin homology  
C;Keywords: hydrolase; metalloproteinase  
F;209-291/Domain: disintegrin homology <DIS>  
F;143/Active site: Glu #status predicted

Query Match 2.9%; Score 7; DB 2; Length 419;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 LFCNVND 83  
|||||

Db 367 LFCNVND 373  
|||||

## RESULT 58

H89860  
conserved hypothetical protein SA0804 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: H89860  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: H89860  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-438 <KUR>  
A;Cross-references: GB:BA000018; PID:G13700747; PIDN:BA842043.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0804  
C;Superfamily: conserved integral membrane protein HP0758

Query Match 2.9%; Score 7; DB 2; Length 438;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDSGSPA 13  
|||||

Db 386 GDSGSPA 392  
|||||

## RESULT 59

T38239

hypothetical protein SPAC23C11.01 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C;Accession: T38239  
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1995  
A;Reference number: Z21781  
A;Accession: T38239  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-441 <BRO>  
A;Cross-references: EMBL:Z98559; PIDN:CA811154.1; GSPDB:GN00066; SPDB:SPAC23C11.01  
A;Experimental source: strain 972h; cosmid c23C11  
C;Genetics:  
A;Gene: SPDB:SPAC23C11.01  
A;Map position: 1  
A;Introns: 88/3; 193/3; 255/3; 293/2

Query Match 2.9%; Score 7; DB 2; Length 441;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TGOALAS 174  
|||||

Db 430 TGOALAS 436  
|||||

## RESULT 60

E70590

3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) - Mycobacterium tuberculosis  
N;Alternate names: aroA protein  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: E70590; A37807  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: E70590  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-450 <COL>  
A;Cross-references: GB:Z95121; GB:AL123456; NID:G3261742; PIDN:CA808328.1; PID:G2072694  
A;Experimental source: strain H37Rv  
R;Garbe, T.; Jones, C.; Charles, I.; Dougan, G.; Young, D.  
J. Bacteriol. 172, 6774-6782, 1990

Query Match 2.9%; Score 7; DB 2; Length 441;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TGOALAS 174  
|||||

Db 430 TGOALAS 436  
|||||

A:Title: Cloning and characterization of the *aroA* gene from *Mycobacterium tuberculosis*.  
A:Reference number: A37807; MUID:91072223; PMID:1212856  
A:Accession: A37807  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-450 <GAR>  
A:Cross-references: GB:M62708; MID:g149927; PIDN:AAA25356.1; PID:g149928  
C:Genetics:  
A:Gene: *aroA*  
C:Superfamily: 3-phosphoshikimate 1-carboxyvinyltransferase; 3-phosphoshikimate 1-carboxyvinyltransferase  
F:15-417/Domain: 3-phosphoshikimate 1-carboxyvinyltransferase homology <PSK>

Query Match 2.9%; Score 7; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 321 ALASPGS 327

RESULT 61  
T30603  
perlecan homolog 2L - Molluscum contagiosum virus 1  
N:Alternate names: MC002L  
C:Species: Molluscum contagiosum virus 1  
C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-May-2000  
C:Accession: T30603  
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.  
Science 273, 813-816, 1996  
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re  
A:Reference number: Z20876; MUID:96325459; PMID:8670425  
A:Accession: T30603  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-451 <SEN>  
A:Cross-references: EMBL:U60315; PIDN:AAC55130.1  
C:Genetics:  
A:Note: MC002L

Query Match 2.9%; Score 7; DB 2; Length 451;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 TSAGSEG 167  
Db 156 TSAGSEG 162

RESULT 62  
S18804  
collagen alpha 4(IV) chain - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 08-Nov-1996  
C:Accession: S18804; S20834; E35167; S20673; D39419; A56630; S19076  
R:Maruyama, M.; Kalluri, R.; Hudson, B.G.; Reeders, S.T.  
J. Biol. Chem. 267, 1253-1258, 1992  
A:Title: The alpha 4(IV) chain of basement membrane collagen. Isolation of cDNAs encodin  
A:Reference number: S18804; MUID:92112769; PMID:1370461  
A:Accession: S18804  
A:Molecule type: mRNA  
A:Residues: 1-453 <MAR>  
A:Cross-references: GB:M77480  
A:Accession: S20834  
A:Molecule type: protein  
A:Residues: 317,'A',319-322,'S',324-328 <MA2>  
R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5466-5469, 1990  
A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type  
A:Reference number: A35167; MUID:90202779; PMID:2318822  
A:Accession: B35167  
A:Molecule type: protein

A:Residues: 217-218,'P',220-242,'X',244-246 <GUN>  
R:Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane c  
A:Reference number: S18432; MUID:87222419; PMID:2430283  
A:Accession: S20673  
A:Molecule type: protein  
A:Residues: 217-218,'P',220-233 <BUT>  
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe  
J. Biol. Chem. 266, 15318-15324, 1991  
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll.  
A:Reference number: A39419; MUID:91332055; PMID:1869555  
A:Accession: D39419  
A:Molecule type: protein  
A:Residues: 217-237 <GU2>  
R:Matsumura, H.; Michael, A.F.; Fish, A.J.; Butkowski, R.J.  
Connect. Tissue Res. 28, 231-244, 1992  
A:Title: Partial protein sequence of the globular domain of alpha 4(IV) collagen chain:  
A:Reference number: A56630; MUID:93105615; PMID:1468209  
A:Accession: A56630  
A:Molecule type: protein  
A:Residues: 217-218,'P',220-242,'X',244-256;258-275,'X',277-278;303-314;391-396,'X',398-  
A:Experimental source: kidney, basement membrane  
A:Note: sequence modified after extraction from NCBI backbone  
C:Species: Thermotoga maritima (strain MSB8)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 23-Dec-2002  
C:Accession: D72344  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Query Match 2.9%; Score 7; DB 2; Length 453;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPGSCL 180  
Db 381 SPGSCL 387

RESULT 63  
D72344  
DNA polymerase III, gamma and tau subunit - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 23-Dec-2002  
C:Accession: D72344  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Query Match 2.9%; Score 7; DB 2; Length 478;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 232 IISRCQV 238  
Db 164 IISRCQV 170

RESULT 64

HMIVS2  
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)  
C:Species: influenza A virus  
C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: A29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.  
A:Reference number: A94370; MUID:88101364; PMID:3336940  
A:Accession: A29971  
A:Molecule type: Genomic RNA  
A:Cross-references: GB:M19056; NID:G324208  
A:Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:G324209) differs from B  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted  
Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 358 SEGTGQA 364  
RESULT 65  
HMIVS3  
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)  
C:Species: influenza A virus  
C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Sep-1998  
C:Accession: B29971  
R:Kida, H.; Shortridge, K.F.; Webster, R.G.  
Virology 162, 160-166, 1988  
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.  
A:Reference number: A94370; MUID:88101364; PMID:3336940  
A:Accession: B29971  
A:Molecule type: Genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M19057; NID:G324210  
A:Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:G324211) differs from B  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted  
Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 358 SEGTGQA 364  
RESULT 66  
HMIV77  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999

C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: A27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: A27813  
A:Molecule type: Genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16737; NID:G324081; PID:AAA43143.1; PID:G324082  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted  
Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 358 SEGTGQA 364  
RESULT 67  
HMIV80  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: B27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: B27813  
A:Molecule type: Genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16738; NID:G324083  
A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted  
Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 358 SEGTGQA 364  
RESULT 68  
HMIV33  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999



C:Accession: C27813  
R:Kida, H.; Kawakata, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: C27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: palmitate (Cys) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

RESULT 69  
HMIV98  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: D27813  
R:Kida, H.; Kawakata, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: D27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: palmitate (Cys) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

RESULT 70  
HMIV21  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: D27813  
R:Kida, H.; Kawakata, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: E27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16741; NID:g324089  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: palmitate (Cys) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

RESULT 71  
HMIV98  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: F27813  
R:Kida, H.; Kawakata, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: F27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16742; NID:g324091  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:520-536/Domain: transmembrane #status predicted <TM1>  
F:8,22,38,165,285,483/Binding site: palmitate (Cys) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

RESULT 72  
HMIV15  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999  
C:Accession: G27813  
R:Kida, H.; Kawakata, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458; PMID:2440178  
A:Accession: G27813  
A:Molecule type: genomic RNA

A;Residues: 1-550 <KID>  
A;Cross-references: GB:M16743; NID:G324093; PIDN:AAA43149.1; PID:G324094  
C;Genetics:  
A;Map position: segment 4  
C;Superfamily: influenza virus hemagglutinin  
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F;520-536/Domain: transmembrane #status predicted <TM1>  
F;18,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

## RESULT 73

EMV86  
hemagglutinin precursor - influenza A virus (strain A/Mem/6/66 [H3N2]) (fragment)  
N;Contains: hemagglutinin HA1; hemagglutinin HA2  
C;Species: influenza A virus  
C;Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jul-1999  
C;Accession: A29245

R;Yasuda, J.M.; Webster, R.G.  
Virology 165, 446-456, 1988  
A;Title: Antigenic and structural characterization of multiple subpopulations of H3N2 in  
A;Reference number: A29245; MUID:88306236; PMID:3407150  
A;Accession: A29245  
A;Molecule type: genomic RNA  
A;Residues: 1-550 <KAT>  
A;Cross-references: GB:M21649; NID:G324295; PIDN:AAA43275.1; PID:G324296  
C;Genetics:  
A;Gene: HA

A;Map position: segment 4  
C;Superfamily: influenza virus hemagglutinin  
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F;520-536/Domain: transmembrane #status predicted <TM1>  
F;18,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

## RESULT 74

JQ1153  
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)  
N;Contains: hemagglutinin HA1; hemagglutinin HA2  
C;Species: influenza A virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C;Accession: JQ1153  
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3  
A;Reference number: JQ1153; MUID:91341491; PMID:1875195  
A;Accession: JQ1153  
A;Molecule type: genomic RNA  
A;Residues: 1-550 <YAS>  
A;Cross-references: GB:D00929; NID:G221279; PIDN:BAA00769.1; PID:G221280

A;Note: the authors translated the codon GGG for residue 218 as Glu  
A;Note: residues 528-532 are not shown in this publication  
C;Superfamily: influenza virus hemagglutinin  
C;Keywords: glycoprotein; homotrimer  
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

## RESULT 75

JQ1154  
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)  
N;Contains: hemagglutinin HA1; hemagglutinin HA2  
C;Species: influenza A virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C;Accession: JQ1154  
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H  
A;Reference number: JQ1153; MUID:91341491; PMID:1875195

A;Accession: JQ1154  
A;Molecule type: genomic RNA  
A;Residues: 1-550 <YAS>  
A;Cross-references: GB:D00930; NID:G221273; PIDN:BAA00770.1; PID:G221274  
A;Note: the authors translated the codon GGG for residue 218 as Glu  
A;Note: residues 528-532 are not shown in this publication  
C;Superfamily: influenza virus hemagglutinin  
C;Keywords: glycoprotein; homotrimer  
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
Db 358 SEGTGQA 364

## RESULT 76

JQ1155  
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)  
N;Contains: hemagglutinin HA1; hemagglutinin HA2  
C;Species: influenza A virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000  
C;Accession: JQ1155  
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.  
J. Gen. Virol. 72, 2007-2010, 1991

A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H  
A;Reference number: JQ1153; MUID:91341491; PMID:1875195

A;Accession: JQ1155  
A;Molecule type: genomic RNA  
A;Residues: 1-550 <YAS>

A;Cross-references: GB:D00931; NID:G221277; PIDN:BAA00771.1; PID:G221278  
A;Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 53  
A;Note: residues 528-532 are not shown in this publication  
C;Superfamily: influenza virus hemagglutinin  
C;Keywords: glycoprotein; homotrimer  
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>  
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>  
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;

```

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
    |||||
Db 358 SEGTGOA 364

RESULT 77
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/231/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01156
R:Yasuda, J.; Shorridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491; PMID:1875195
A:Accession: J01156
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00932; NID:9221275; PIDN:BA00772.1; PID:9221276
A>Note: the authors translated the codon TCG for residue 215 as Pro and GGA for residue
A:Note: residues 528-532 are not shown in this publication
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
    |||||
Db 358 SEGTGOA 364

RESULT 78
HMIVE1
hemagglutinin precursor - influenza A virus (strain A/equine/Uruguay/1/63 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: A34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: A34064
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24718; GB:J04336; NID:9324024; PIDN:AAA43114.1; PID:9324025
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
    |||||
Db 358 SEGTGOA 364

```

```

Db 373 SEGTGOA 379

RESULT 79
HMIVE3
hemagglutinin precursor - influenza A virus (strain A/equine/Tokyo/71 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: C34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: C34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24720; GB:J04336; NID:9324018; PIDN:AAA43111.1; PID:9324019
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
    |||||
Db 373 SEGTGOA 379

RESULT 80
HMIVE4
hemagglutinin precursor - influenza A virus (strain A/equine/Algiers/72 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: D34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: D34064
A:Molecule type: Genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24721; GB:J04336; NID:9323996; PIDN:AAA43100.1; PID:9323997
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
    |||||
Db 373 SEGTGOA 379

```

```
RESULT 81
HMIVE6
hemagglutinin precursor - influenza A virus (strain A/equine/New Market/76 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: F34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: F34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24722; GB:J04336; NID:g324010; PIDN:AAA43107.1; PID:g324011
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:18,23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379

RESULT 82
HMIVE6
hemagglutinin precursor - influenza A virus (strain A/equine/Fontainebleau/76 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: F34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: F34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24723; GB:J04336; NID:g323998; PIDN:AAA43101.1; PID:g323999
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:18,23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379
```

```
RESULT 83
HMIVE7
hemagglutinin precursor - influenza A virus (strain A/equine/Romania/80 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: G34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: G34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24724; GB:J04336; NID:g324014; PIDN:AAA43109.1; PID:g324015
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:18,23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379

RESULT 84
HMIVE8
hemagglutinin precursor - influenza A virus (strain A/equine/Santiago/1/85 [H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: H34064
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A:Reference number: A34064; MUID:89204899; PMID:2705299
A:Accession: H34064
A:Molecule type: genomic RNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:M24725; GB:J04336; NID:g324016; PIDN:AAA43110.1; PID:g324017
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:535-551/Domain: transmembrane #status predicted <TM1>
F:18,23,37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379
```

## RESULT 85

HMIVF5  
hemagglutinin precursor - influenza A virus (strain A/equine/Tennessee/5/85 [H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: J34064  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virolgy 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; MUID:89204899; PMID:2705299  
A:Accession: J34064  
A:Molecule type: Genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24726; GB:J04336; NID:g324020; PIDN:AAA43112.1; PID:g324021  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>  
F:535-551/Domain: transmembrane #status predicted <TM1>  
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:129-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted  
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 373 SEGTGQA 379

## RESULT 86

HMIVET  
hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/2/86 [H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: A34065  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virolgy 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; MUID:89204899; PMID:2705299  
A:Accession: A34065  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24727; GB:J04336; NID:g324000; PIDN:AAA43102.1; PID:g324001  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>  
F:535-551/Domain: transmembrane #status predicted <TM1>  
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:129-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted  
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 373 SEGTGQA 379

## RESULT 87

## HMIVF6

hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/1/87 [H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C>Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: B34065  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virolgy 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; MUID:89204899; PMID:2705299  
A:Accession: B34065  
A:Molecule type: Genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24728; GB:J04336; NID:g324002; PIDN:AAA43103.1; PID:g324003  
C:Genetics:

A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>  
F:535-551/Domain: transmembrane #status predicted <TM1>  
F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:129-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted  
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 565;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 373 SEGTGQA 379

## RESULT 88

S33703  
hemagglutinin - influenza A virus H3N8  
C:Species: influenza A virus H3N8, equine influenza virus  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999  
C:Accession: S33703  
R:Binns, M.M.; Daly, J.M.; Chirnside, E.D.; Mumford, J.A.; Wood, J.M.; Richards, C.M.; E  
Arch. Virol. 130, 33-43, 1993  
A:Title: Genetic and antigenic analysis of an equine influenza H3 isolate from the 1989  
A:Reference number: S33703; MUID:93277383; PMID:8503788  
A:Accession: S33703  
A>Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <BIN>  
A:Cross-references: EMBL:X68437; NID:g312668; PIDN:CAA48482.1; PID:g312669  
A>Note: the authors translated the codon ACC for residue 403 as Arg  
C:Superfamily: influenza virus hemagglutinin

Query Match 2.9%; Score 7; DB 2; Length 565;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 373 SEGTGQA 379

## RESULT 89

HMIVH  
hemagglutinin precursor - influenza A virus  
C:Species: influenza A virus  
C>Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 22-Oct-1999  
C:Accession: A93705; A93233; A04051; A93231; A94441  
R:Both, G.W.; Sleight, M.J.  
Nucleic Acids Res. 8, 2561-2575, 1980  
A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza  
A:Reference number: A93705; MUID:81053698; PMID:6253883  
A:Accession: A93705

A:Molecule type: genomic RNA  
A:Residues: 1-566 <BOT>  
A:Cross-references: GB:V01103  
A:Experimental source: strain A/NT/60/68/29C  
A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/68/29C  
R:Dopheide, T.A.; Ward, C.W.  
FEBS Lett. 110, 181-183, 1980  
A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.  
A:Reference number: A91276; MUID:8019105; PMID:6768586  
A:Contents: annotation; disulfide bonds  
R:Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.  
Nature 287, 301-306, 1980  
A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from A/Reference number: A93233; MUID:81030852; PMID:7421990  
A:Accession: A93233  
A:Molecule type: genomic RNA  
A:Residues: 1-24, 'S', '26', 'D', '28-159, 'G', '161-197, 'I', '199-241, 'L', '243-249 <GET>  
A:Experimental source: strain X-31[H3]  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:536-552/Domain: transmembrane #status predicted <TM1>  
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status experimental  
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted  
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 374 SEGTGQA 380

RESULT 90  
HMIVHA  
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
A:Accession: A93231; A04051  
R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.  
Nature 286, 171-176, 1980  
A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Reference number: A93231; MUID:80254693; PMID:7402351  
A:Accession: A93231  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <VAR>  
A:Cross-references: GB:J02090; NID:G324131; PIDN:AAA43178.1; PID:G324132  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 374 SEGTGQA 380

RESULT 91  
HMIVHM  
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 31-Mar-2000

C:Accession: A94441; A04051  
R:Sleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.  
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, 1977  
A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of C:Accession: A94441  
A:Reference number: A94441  
A:Accession: A94441  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <SLB>  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 374 SEGTGQA 380

RESULT 92  
HMIIV6  
hemagglutinin precursor - influenza A virus (strain A/England/321/77)  
C:Species: influenza A virus  
C:Date: 05-Apr-1983 #sequence\_revision 14-Nov-1983 #text\_change 22-Oct-1999  
A:Accession: B92790; A92979; A04052  
R:Hauptmann, R.; Clarke, L.D.; Mountford, R.C.; Bachmayer, H.; Almond, J.W.  
J. Gen. Virol. 64, 215-220, 1983  
A:Title: Nucleotide sequence of the haemagglutinin gene of influenza virus A/England/321/77  
A:Reference number: A92790; MUID:83110955; PMID:6822816  
A:Accession: B92790  
A:Molecule type: genomic RNA  
A:Residues: 1-566 <HAU>  
A:Cross-references: GB:X05907; NID:G60694; PIDN:CAA29337.1; PID:G60695  
A:Note: the authors translated the codon GUU for residue 14 as Asn, GCC for residue 16 as Arg.  
R:Both, G.W.; Sleight, M.J.  
J. Virol. 39, 663-672, 1981  
A:Title: Conservation and variation in the hemagglutinins of Hong Kong subtype influenza A/Reference number: A92979; MUID:82033259; PMID:6169840  
A:Accession: A92979  
A:Molecule type: genomic RNA  
A:Residues: 17-31, 'G', '33-148, 'S', '150-158, 'S', '160-171, 'E', '173-175, 'K', '177-187, 'G', '189-208  
A:Cross-references: GB:J02092; NID:G324139; PIDN:AAA43182.1; PID:G324140  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F:536-552/Domain: transmembrane #status predicted <TM1>  
F:30-482, 68-293, 80-92, 155-489, 297-321/Disulfide bonds: #status predicted  
F:555, 562, 565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||  
DB 374 SEGTGQA 380

RESULT 93  
HMIIVU  
hemagglutinin precursor - influenza A virus (strain A/duck/Ukraine/63)  
C:Species: influenza A virus  
C:Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 28-May-1999  
C:Accession: A04053

R;Fang, R.; Min Jou, W.; Huylebroeck, D.; Devos, R.; Fiers, W.  
Cell 25, 315-323, 1981  
A;Title: Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal vi  
A;Reference number: A04053; MUID:82025542; PMID:6169439  
A;Accession: A04053  
A;Molecule type: genomic RNA  
A;Residues: 1-566 <PAN>  
A;Cross-references: GB:J02109; GB:J02108; NID:G60756; PIDN:CAA24271.1; PID:G60757  
C;Superfamily: Influenza virus hemagglutinin  
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>  
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>  
F;536-552/Domain: transmembrane #status predicted <TM1>  
F;530-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status predicted  
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
| | | | |  
Db 374 SEGTGQA 380

RESULT 94  
HMIVV  
hemagglutinin precursor - influenza A virus (strain A/Victoria/3/75)  
C;Species: Influenza A virus  
C;Date: 28-Feb-1981 #sequence\_revision 28-Feb-1981 #text\_change 28-May-1999  
C;Accession: A07994; A04050; A92790  
C;Min Jou, W.; Verhoeven, M.; Devos, R.; Saman, E.; Fang, R.; Huylebroeck, D.; Fiers, W.  
Cell 19, 683-696, 1980  
A;Title: Complete structure of the hemagglutinin gene from the human influenza A/Victoria  
A;Reference number: A90794; MUID:80155186; PMID:6153390  
A;Accession: A90794  
A;Molecule type: genomic RNA  
A;Residues: 1-567 <MIN>  
A;Cross-references: GB:V01098; GB:J02172; GB:M55060; NID:G60784; PIDN:CAA24281.1; PID:G6  
C;Genetics:  
A;Map position: segment 4  
C;Superfamily: Influenza virus hemagglutinin  
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-345/Product: hemagglutinin HA1 #status predicted <HA1>  
F;347-567/Product: hemagglutinin HA2 #status predicted <HA2>  
F;537-553/Domain: transmembrane #status predicted <TM1>  
F;31-483,69-294,81-93,156-490,298-322/Disulfide bonds: #status predicted  
F;556,563,566/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.9%; Score 7; DB 1; Length 567;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
| | | | |  
Db 375 SEGTGQA 381

RESULT 95  
A45137  
collagen alpha 4(IV) chain - rabbit  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-1999  
C;Accession: A45137  
R;Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of  
A;Reference number: S28777; MUID:93054733; PMID:1429714  
A;Accession: A45137  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA

A;Residues: 1-623 <KAM>  
A;Experimental source: basement membrane  
A;Note: sequence extracted from NCBI backbone (NCBIP:118549)  
C;Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPQSCLE 180  
| | | | |  
Db 551 SPQSCLE 557

RESULT 96  
A30347  
exotoxin A precursor - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa  
C;Date: 08-Jun-1990 #sequence\_revision 08-Jun-1990 #text\_change 24-Nov-1999  
C;Accession: A30347  
R;Gray, G.L.; Smith, D.H.; Baldridge, J.S.; Harkins, R.N.; Vasil, M.L.; Chen, E.Y.; Heyt  
Proc. Natl. Acad. Sci. U.S.A. 81, 2645-2649, 1984  
A;Title: Cloning, nucleotide sequence, and expression in Escherichia coli of the exotoxi  
A;Reference number: A30347; MUID:84194063; PMID:6201861  
A;Accession: A30347  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-638 <GRA>  
A;Cross-references: GB:K01397; GB:M23348; NID:G151215; PIDN:AAB59097.1; PID:G151216  
C;Superfamily: Pseudomonas aeruginosa exotoxin A  
C;Keywords: exotoxin

Query Match 2.9%; Score 7; DB 2; Length 638;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
| | | | |  
Db 340 ALASPGS 346

RESULT 97  
C83503  
exotoxin A precursor PA1148 [imported] - Pseudomonas aeruginosa (strain PA01)  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: C83503  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B:  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: C83503  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-638 <STO>  
A;Cross-references: GB:AEC04544; GB:AEC04091; NID:G9947060; PIDN:AA04537.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: toxA; PA1148  
C;Superfamily: Pseudomonas aeruginosa exotoxin A

Query Match 2.9%; Score 7; DB 2; Length 638;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
| | | | |  
Db 340 ALASPGS 346

RESULT 98



I38755  
transcription factor REST (version 2) - human (fragment)  
N:Alternate names: neural-restrictive silencer factor; RE1-silencing transcription factor  
C:Species: Homo sapiens (man)  
C>Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 05-Nov-1999  
C:Accession: I38755  
R:Schonherr, C.J.; Anderson, D.J.  
Science 267, 1360-1363, 1995  
A:Title: The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple genes  
A:Reference number: I38754; MUID:95176234; PMID:7871435  
A:Accession: I38755  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-681 <RES>  
A:Cross-references: EMBL:U13879; NID:G606947; PIDN:AAC50115.1; PID:G606948  
C:Genetics:  
A:Gene: GDB:REST; NRSF  
A:Cross-references: GDB:702138  
A:Map position: 4q12-4q12  
C:Keywords: transcription regulation

Query Match 2.9%; Score 7; DB 2; Length 681;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSPA 13  
|||||  
Db 66 GDSGSPA 72

RESULT 99  
S49228  
sodium-dependent phosphate transporter - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 05-Nov-1999  
C:Accession: S68972; S49228  
R:Helms, C.; Murer, H.; McGivan, J.  
Eur. J. Biochem. 228, 927-930, 1995  
A:Title: Cloning, sequence analysis and expression of the cDNA encoding a sodium-dependent phosphate transporter  
A:Reference number: S68972; MUID:95255303; PMID:7737195  
A:Accession: S68972  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-693 <HE2>  
A:Cross-references: EMBL:X61699; NID:G547483; PIDN:CAA57345.1; PID:G547484

Query Match 2.9%; Score 7; DB 2; Length 693;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
|||||  
Db 472 ALASPGS 478

RESULT 100  
AC0018  
probable membrane protein YPO0142 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AC0018  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.G.  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0018  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-715 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89005.1; PID:G15978247; GSPDB:GN00175  
C:Genetics:

A:Gene: YPO0142  
C:Superfamily: Escherichia coli hypothetical protein yrfF

Query Match 2.9%; Score 7; DB 2; Length 715;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TRGFVFT 23  
|||||  
Db 422 TRGFVFT 428

RESULT 101  
T09395  
envelope polyprotein - walleye dermal sarcoma virus  
C:Species: walleye dermal sarcoma virus  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999  
C:Accession: T09395  
R:Petroopoulos, C.J.  
submitted to the EMBL Data Library, November 1997  
A:Description: Appendix 2: Retroviral taxonomy, protein structure, sequences, and genetics  
A:Reference number: Z16660  
A:Accession: T09395  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: Genomic RNA  
A:Residues: 1-1225 <PEI>  
A:Cross-references: EMBL:AF033822; NID:G2801519; PID:G2801522  
C:Genetics:  
A:Gene: env  
C:Superfamily: walleye dermal sarcoma virus envelope polyprotein

Query Match 2.9%; Score 7; DB 2; Length 1225;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 PMNWAPI 110  
|||||  
Db 967 PMNWAPI 973

RESULT 102  
T13123  
DNA replication primase protein - phase N15  
N:Alternate names: protein gp37; replication protein  
C:Species: phase N15  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: T13123; T13285  
R:Hendrix, R.W.; Ravin, V.K.; Casjens, S.R.; Ford, M.E.; Ravin, N.V.; Smirnov, I.K.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z17603  
A:Accession: T13123  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1324 <HEN>  
A:Cross-references: EMBL:AF064539; NID:G3192683; PID:G3192713; PIDN:AAC19066.1  
R:Vostrikhina, O.A.; Vostrov, A.A.; Rybchin, V.N.; Svarchevsky, A.N.  
submitted to the EMBL Data Library, July 1996  
A:Description: Characterization of the repA region of the Escherichia coli linear plasmid pRepA  
A:Reference number: Z17650  
A:Accession: T13285  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1207; VYQFTHVRWK, 1218, 'P', 1264, 'SK', 1267, 'TAIW', <VOS>  
A:Cross-references: EMBL:U63085; NID:G2529379; PID:G2529380; PIDN:AAC48876.1  
A:Experimental source: specific\_host Escherichia coli  
C:Genetics:  
A:Gene: repA  
A>Note: Gene 37

Query Match 2.9%; Score 7; DB 2; Length 1324;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SGFSFLF 49

Db 105 SGFSFLF 111

## RESULT 103

T00333

Hypothetical protein KIAA0560 - human

C/Species: Homo sapiens (man)

C/Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 11-Jan-2002

C/Accession: T00333

R/Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A/Title: Prediction of the coding sequences of unidentified human genes. IX. The completed

A/Reference number: Z14086; MUID:98290545; PMID:9628581

A/Accession: T00333

A/Status: preliminary; translated from GB/ENBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1421 &lt;NAG&gt;

A/Cross-references: EMBL:AB011132; NID:dl185402; PIDN:BAA25486.1

A/Experimental source: brain; clone HH1648

C/Genetics:

A/Note: KIAA0560

## Query Match

Best Local Similarity 2.9%; Score 7; DB 2; Length 1421;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 LEEPRAS 185

Db 944 LEEPRAS 950

## RESULT 104

F86366

protein F26F24.8 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: F86366

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C/A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: F86366

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1583 &lt;STO&gt;

A/Cross-references: GB:AB005172; NID:99295691; PIDN:AAF86997.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

## Query Match

Best Local Similarity 2.9%; Score 7; DB 2; Length 1583;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRGD 8

Db 773 LKGRGD 779

## RESULT 105

CGH11B

collagen alpha 4(IV) chain precursor - human

N/Alternate names: procollagen alpha 4(IV) chain

C/Species: Homo sapiens (man)

C/Date: 06-Feb-1995 #sequence\_revision 03-Oct-1995 #text\_change 16-Jun-2000  
 C/Accession: A55360; S36854; S28777  
 R/Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.  
 J. Biol. Chem. 269, 26172-26177, 1994  
 A/Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp  
 A/Reference number: A55360; MUID:95014445; PMID:7523402  
 A/Accession: A55360

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1690 &lt;LEI&gt;

A/Cross-references: GB:X81053; NID:9574805; PIDN:CAA56943.1; PID:9574806

R/Sugimoto, M.; Ohashi, T.; Yoshiohka, H.; Matsuo, N.; Ninomiya, Y.

FEBS Lett. 330, 122-128, 1993

A/Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha

A/Reference number: S36854; MUID:93374047; PMID:8365481

A/Accession: S36854

A/Molecule type: DNA; mRNA

A/Residues: 1219-1658, 'PE', 1661-1690 &lt;SUG&gt;

A/Cross-references: DDBJ:DL7391; NID:9440365; PIDN:BAA04214.1; PID:9457161

A/Experimental source: whole eye

R/Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.

J. Biol. Chem. 267, 23753-23758, 1992

A/Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of

A/Reference number: S28777; MUID:93054733; PMID:1429714

A/Accession: S28777

A/Molecule type: DNA

A/Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 &lt;KAM&gt;

A/Cross-references: GB:L01475; GB:L01476

A/Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the

C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

ed and subsequently O-glycosylated.

C/Genetics:

A/Gene: GDB:COL4A4

A/Cross-references: GDB:L32673; OMIM:120131

A/Map position: 2q35-2q37

A/Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete

A/Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with

C/Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3(

among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a

C/Associations: in the interrupted helical domain (with disulfide and desmosine cross-li

C/Function:

A/Description: minor structural component of extracellular basement membrane in kidney s

C/Superfamily: collagen alpha 1(IV) chain

C/Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxy

F/1-38/Domain: signal sequence #status predicted &lt;SIG&gt;

F/39-1690/Product: collagen alpha 4(IV) chain #status predicted &lt;MAT&gt;

F/39-61/Domain: amino-terminal nonhelical, NH1 &lt;NH1&gt;

F/62-1466/Region: interrupted helical

F/94-96/Region: cell attachment (R-G-D) motif

F/145-147/Region: cell attachment (R-G-D) motif

F/189-191/Region: cell attachment (R-G-D) motif

F/310-312/Region: cell attachment (R-G-D) motif

F/724-726/Region: cell attachment (R-G-D) motif

F/785-787/Region: cell attachment (R-G-D) motif

F/989-991/Region: cell attachment (R-G-D) motif

F/1212-1214/Region: cell attachment (R-G-D) motif

F/1467-1690/Domain: carboxyl-terminal nonhelical, NC1 &lt;NC1&gt;

F/1471-1569/Domain: collagen IV carboxyl-terminal repeat &lt;CT1&gt;

F/1579-1686/Domain: collagen IV carboxyl-terminal repeat &lt;CT2&gt;

F/47,82,85,87,286,400,460,492,494,668,790,828,1095,1131,1294,1317,1375,1407/Disulfide bc

F/1480-1566/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/1525-1531,1634-1641/Disulfide bonds: #status predicted

F/1588-1683,1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

## Query Match

Best Local Similarity 2.9%; Score 7; DB 1; Length 1690;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180

Db 1618 SPGSCLE 1624

## RESULT 106

T29350  
Hypothetical protein F01G12.5a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T29350  
R:Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid F01G12.5a  
A:Reference number: Z20611  
A:Accession: T29350  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1758 <WUX>  
A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a  
A:Experimental source: strain Bristol N2; clone F01G12  
C:Genetics:  
A:Gene: CESP:F01G12.5a  
A:Map position: X  
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 1758;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLORF 71

|||||

DB 1577 GSCLORF 1583

## RESULT 107

T29351  
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans  
N:Alternate names: collagen alpha 2(IV) chain precursor clb-1  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T29351  
R:Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid F01G12.5a  
A:Reference number: Z20611  
A:Accession: T29351  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1759 <WUX>  
A:Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a  
A:Experimental source: strain Bristol N2; clone F01G12  
C:Genetics:  
A:Gene: CESP:F01G12.5a  
A:Map position: X  
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 2.9%; Score 7; DB 2; Length 1759;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLORF 71

|||||

DB 1578 GSCLORF 1584

## RESULT 108

S16366  
collagen alpha 2(IV) chain precursor - pig roundworm  
C:Species: Ascaris suum (pig roundworm)  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C:Accession: S16366  
R:Pettitt, J.; Kingston, I.B.  
J. Biol. Chem. 266, 16149-16156, 1991  
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the part

A:Reference number: S16366; MUID:91340768; PMID:1714907

A:Accession: S16366

A:Molecule type: mRNA

A:Residues: 1-1763 <JBI>

C:Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649

C:Genetics:

A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfide

F1-26/Domain: signal sequence #status predicted <SIG>

F127-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>

F127-1763/Domain: non-collagenous NHI #status predicted <NHI>

F127-1763/Domain: collagenous #status predicted <COL>

F197-199/Region: cell attachment (R-G-D) motif

F1530-1763/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCL>

F1530-1763/Domain: repeat NCI #status predicted <NCL1>

F1530-1763/Domain: repeat NCI #status predicted <NCL2>

F126/Binding site: carbohydrate (Asn) (covalent) #status predicted

F1593-1599.1702-1709/Disulfide bonds: #status predicted

Query Match 2.9%; Score 7; DB 2; Length 1763;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMPFLFC 79

|||||

DB 1587 TMPFLFC 1593

## RESULT 109

T08991  
hypothetical protein F6G3.180 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C:Accession: T08991  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.;  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16520  
A:Accession: T08991  
A:Molecule type: DNA  
A:Residues: 1-1966 <BEV>  
A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.180  
A:Experimental source: cultivar Columbia; BAC clone F6G3  
C:Genetics:  
A:Gene: ATSP:F6G3.180  
A:Map position: 4  
A:Introns: 113/2; 652/3; 1112/3; 1220/2; 1720/1; 1868/2; 1934/2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F6G3.180

Query Match 2.9%; Score 7; DB 2; Length 1966;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171

|||||

DB 1342 SEGTGQA 1348

## RESULT 110

S02041  
dystrophin, muscle - chicken  
N:Alternate names: duchenne muscular dystrophy protein  
C:Species: Gallus gallus (chicken)  
C>Date: 07-Sep-1990 #sequence\_revision 27-Jun-1994 #text\_change 16-Jul-1999  
C:Accession: S02041; S02013; S71487  
R:Leclaire, C.; Heilig, R.; Mandel, J.L.  
Nucleic Acids Res. 16, 11815-11816, 1988  
A:Title: Nucleotide sequence of chicken dystrophin cDNA.  
A:Reference number: S02041; MUID:89098331; PMID:3062582  
A:Accession: S02041  
A:Status: translation not shown  
A:Molecule type: mRNA

A;Residues: 1-3660 <LEM>  
A;Cross-references: EMBL:X13369, NID:963369; PIDN:CAA31746.1; PID:963370  
A;Note: 1869-His, 1869-Arg, and sequences lacking 1171-Met were also found  
R;Lemaire, C.; Heilig, R.; Mandel, J.L.  
EMBO J. 7, 4157-4162, 1988  
A;Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding and  
A;Reference number: S02013; MUID:89210800; PMID:3072195  
A;Accession: S02013  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-3573, 'HA', 3576-3660 <LEM2>  
R;Heilig, R.; Lemaire, C.; Mandel, J.L.  
Nucleic Acids Res. 15, 9129-9142, 1987  
A;Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection of a cor  
A;Reference number: S09071; MUID:88067745; PMID:2825128  
A;Accession: S71487  
A;Molecule type: DNA  
A;Residues: 222-281 <HE1>  
C;Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl  
C;Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro  
C;Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystro  
C;Keywords: actin binding; calmodulin binding; cytoskeleton; leucine zipper; membrane-as  
F;18-237/Domain: alpha-actinin actin-binding domain homology <ACT>  
F;253-327/Region: hinge  
F;340-449/Domain: spectrin/dystrophin repeat homology <SP01>  
F;450-558/Domain: spectrin/dystrophin repeat homology <SP02>  
F;560-669/Domain: spectrin/dystrophin repeat homology <SP03>  
F;670-719/Region: hinge  
F;720-830/Domain: spectrin/dystrophin repeat homology <SP04>  
F;838-936/Domain: spectrin/dystrophin repeat homology <SP05>  
F;940-1047/Domain: spectrin/dystrophin repeat homology <SP06>  
F;1049-1156/Domain: spectrin/dystrophin repeat homology <SP07>  
F;1158-1265/Domain: spectrin/dystrophin repeat homology <SP08>  
F;1267-1369/Domain: spectrin/dystrophin repeat homology <SP09>  
F;1374-1479/Domain: spectrin/dystrophin repeat homology <SP10>  
F;1480-1570/Domain: spectrin/dystrophin repeat homology <SP11>  
F;1572-1678/Domain: spectrin/dystrophin repeat homology <SP12>  
F;1680-1784/Domain: spectrin/dystrophin repeat homology <SP13>  
F;1787-1877/Domain: spectrin/dystrophin repeat homology <SP14>  
F;1878-1984/Domain: spectrin/dystrophin repeat homology <SP15>  
F;1986-2103/Domain: spectrin/dystrophin repeat homology <SP16>  
F;2105-2211/Domain: spectrin/dystrophin repeat homology <SP17>  
F;2213-2319/Domain: spectrin/dystrophin repeat homology <SP18>  
F;2323-2419/Domain: spectrin/dystrophin repeat homology <SP19>  
F;2420-2467/Region: hinge  
F;2468-2574/Domain: spectrin/dystrophin repeat homology <SP20>  
F;2576-2683/Domain: spectrin/dystrophin repeat homology <SP21>  
F;2685-2799/Domain: spectrin/dystrophin repeat homology <SP22>  
F;2801-2928/Domain: spectrin/dystrophin repeat homology <SP23>  
F;2930-3037/Domain: spectrin/dystrophin repeat homology <SP24>  
F;3038-3075/Region: hinge  
F;3082-3089/Region: WW repeat homology <WW1>  
F;3079-3357/Region: cysteine-rich  
F;3481-3502/Region: leucine zipper motif  
F;3547-3568/Region: leucine zipper motif  
Query Match 2.9%; Score 7; DB 1; Length 3660;  
Best Local Similarity 100.0%; Pred.No. 4.1e+02; Indels 0;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
QY 208 ASLNPER 214  
DB 2284 ASLNPER 2290  
RESULT 111  
E39419  
collagen alpha 5(IV) chain - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 03-Apr-1992 #sequence\_revision 03-Apr-1992 #text\_change 19-Oct-1995  
C;Accession: E39419  
R;Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe  
J. Biol. Chem. 266, 15318-15324, 1991

A;Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll  
A;Reference number: A39419; MUID:91332055; PMID:1869555  
A;Accession: E39419  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-20 <GUN>  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
Query Match 2.5%; Score 6; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred.No. 53; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 23 TRHSQT 28  
DB 15 TRHSQT 20  
RESULT 112  
T01664  
envelope protein - hepatitis C virus (fragment)  
C;Species: hepatitis C virus  
C;Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 17-Nov-2000  
C;Accession: T01664  
R;Caraselli, E.; Cerino, A.; Esposito, G.; Sillini, E.; Mondelli, M.U.; Traboni, C.  
J. Virol. 69, 4407-4412, 1995  
A;Title: Occurrence of antibodies reactive with more than one variant of the putative env  
A;Reference number: Z14388; MUID:95287497; PMID:7539508  
A;Accession: T01664  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-27 <SCA>  
A;Cross-references: EMBL:X79669; NID:92276229; PIDN:CAA56117.1; PID:92276230  
C;Genetics:  
A;Gene: B2/NS1  
C;Superfamily: hepatitis C virus genome polypeptide  
Query Match 2.5%; Score 6; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred.No. 68; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 16 TTRGFV 21  
DB 12 TTRGFV 17  
RESULT 113  
I39969  
outer membrane protein A - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 19-Jul-1995 #sequence\_revision 19-Jul-1996 #text\_change 20-Aug-1999  
C;Accession: I39969  
R;Ikemura, H.; Takagi, H.; Inouye, M.  
J. Biol. Chem. 262, 7859-7864, 1987  
A;Title: Requirement of pro-sequence for the production of active subtilisin E in Escher  
A;Reference number: I39969; MUID:87222417; PMID:3108260  
A;Accession: I39969  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-35 <RES>  
A;Cross-references: GB:JM6639; NID:9143521; PIDN:AAA22743.1; PID:9143522  
A;Experimental source: strain W168, substrain P779  
A;Note: sequence was not translated in the genome sequence, reference number A69580  
C;Genetics:  
A;Gene: ompA  
C;Superfamily: outer membrane protein A  
C;Keywords: membrane protein  
Query Match 2.5%; Score 6; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred.No. 85; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 129 AIAIV 134

```
Db          5 AIAIAV 10
|||||
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:/Reference number: AB0502; MUID:21534947; PMID:11677608
A:/Accession: AF0716
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-49 <PAR>
A:/Cross-references: GB:AL513382; PIDN:CAD02103.1; PID:gl6502938; GSPDB:GN00176
C:/Genetics:
A:/Gene: STY1870

RESULT 114
PH1753
IG heavy chain V region (clone NP-12-19) - mouse (fragment)
C:/Species: Mus musculus (house mouse)
C:/Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C:/Accession: PH1753
R:/McHeyzer-Williams, M.G.; McLean, M.J.; Lalor, P.A.; Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A:/Title: Antigen-driven B cell differentiation in vivo.
A:/Reference number: PH1675; MUID:93301607; PMID:8315385
A:/Accession: PH1753
A:/Molecule type: mRNA
A:/Residues: 1-36 <MCH>
A:/Experimental source: B cell
C:/Superfamily: immunoglobulin V region; immunoglobulin homology
C:/Keywords: heterotrimer; immunoglobulin

Query Match          2.5%; Score 6; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 YYSNSY 203
|||||
Db 27 YYSNSY 32

RESULT 115
AD1753
Lactococcus lactis prophage pi2 protein 41 homolog lin2569 [imported] - Listeria innocua
C:/Species: Listeria innocua
C:/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:/Accession: AD1753
R:/Glaser, P.; Frangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fshini, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:/Title: Comparative genomics of Listeria species
A:/Reference number: AB1077; MUID:21537279; PMID:11679669
A:/Accession: AD1753
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-44 <GLA>
A:/Cross-references: GB:AL592022; PIDN:CAC97796.1; PID:gl6415091; GSPDB:GN00178
A:/Experimental source: strain Clp11262
C:/Genetics:
A:/Gene: lin2569

Query Match          2.5%; Score 6; DB 2; Length 44;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 63 TLGSCL 68
|||||
Db 10 TLGSCL 15

RESULT 116
AF0716
hypothetical protein STY1870 [imported] - Salmonella enterica subsp. enterica serovar Ty
C:/Species: Salmonella enterica subsp. enterica serovar Typhi
A:/Note: this species has also been called Salmonella typhi.
C:/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:/Accession: AF0716
R:/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.N.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

Query Match          2.5%; Score 6; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 230 BKTIISR 235
|||||
```

```
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:/Reference number: AB0502; MUID:21534947; PMID:11677608
A:/Accession: AF0716
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-49 <PAR>
A:/Cross-references: GB:AL513382; PIDN:CAD02103.1; PID:gl6502938; GSPDB:GN00176
C:/Genetics:
A:/Gene: STY1870

Query Match          2.5%; Score 6; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 APITGR 113
|||||
Db 15 APITGR 20

RESULT 117
S61509
light-harvesting protein B800-820 alpha chain - Rhodospirillum mollischanum (DSM 119)
C:/Species: Rhodospirillum mollischanum
A:/Variety: DSM 119
C:/Date: 15-Jul-1995 #sequence_revision 01-Mar-1996 #text_change 09-Mar-1996
C:/Accession: S61509
R:/Sauer, P.
submitted to the Protein Sequence Database, February 1996
A:/Description: Deletion of a B800-850 light-harvesting complex in Rhodospirillum molis
A:/Reference number: S61509
A:/Accession: S61509
A:/Molecule type: protein
A:/Residues: 1-56 <SAU>
A:/Experimental source: DSM 119

Query Match          2.5%; Score 6; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 AIAVHS 136
|||||
Db 30 AIAVHS 35

RESULT 118
S75543
hypothetical protein sst1375 - Synechocystis sp. (strain PCC 6803)
C:/Species: Synechocystis sp.
A:/Variety: PCC 6803
C:/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:/Accession: S75543
R:/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:/Reference number: S74322; MUID:97061201; PMID:8905231
A:/Accession: S75543
A:/Status: nucleic acid sequence not shown; translation not shown
A:/Molecule type: DNA
A:/Residues: 1-64 <KAN>
A:/Cross-references: EMBL:D90911; GB:AB001339; NID:gl653083; PIDN:BAAL8104.1; PID:dl01863
A:/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:/Genetics:
A:/Start codon: GTG

Query Match          2.5%; Score 6; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 230 BKTIISR 235
|||||
```

Db 56 EKIISR 61

## RESULT 119

hypothetical protein 65 - Cyanophora paradoxa cyanelle  
N/Alternate names: Cyanophora paradoxa ycf9  
C/Species: Cyanelle Cyanophora paradoxa  
C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 21-Jan-2000  
C/Accession: S14712; T06892; S10372  
R/Erzard, J.L.; Weil, J.H.; Kuntz, M.  
Plant Mol. Biol. 15, 779-781, 1990  
A/Title: An ORF potentially encoding a 6.5 kDa hydrophobic protein in chloroplasts is a  
A/Reference number: S14712; MUID:91346714; PMID:2129400  
A/Accession: S14712  
A/Molecule type: DNA  
A/Residues: 1-65 <EVR>  
A/Cross-references: EMBL:X51491; NID:g12550; PIDN:CAA35786.1; PID:g12551  
R/Schirwail, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.  
submitted to the EMBL Data Library, July 1995  
A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.  
A/Reference number: Z15840  
A/Accession: T06892  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-65 <STI>  
A/Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81235.1; PID:g1016148  
A/Experimental source: strain Pringsheim LB555  
C/Genetics:  
A/Gene: ycf9  
A/Genome: cyanelle  
C/Superfamily: conserved hypothetical protein ycf9  
C/Keywords: cyanelle

Query Match 2.5%; Score 6; DB 2; Length 65;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPG 176

Db 26 ALASPG 31

## RESULT 120

hypothetical protein T29H11.30 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
C/Accession: T06697  
R/Quetier, F.; Choise, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artigou  
submitted to the Protein Sequence Database, April 1999  
A/Reference number: Z15793  
A/Accession: T06697  
A/Molecule type: DNA  
A/Residues: 1-66 <QUR>  
A/Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.30  
A/Experimental source: cultivar Columbia; BAC clone T29H11  
C/Genetics:  
A/Gene: ATSP:T29H11.30  
A/Map position: 3  
A/Introns: 52/1

Query Match 2.5%; Score 6; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 TRHSQT 28

Db 33 TRHSQT 38

## RESULT 121

AB3399

hypothetical protein BMEI1176 [imported] - Brucella melitensis (strain 16M)  
C/Species: Brucella melitensis  
C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C/Accession: AB3399  
R/DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
M.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A/Reference number: AD3252; PMID:11756688  
A/Accession: AB3399  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-72 <KUR>  
A/Cross-references: GB:AE008917; PIDN:AAL52357.1; PID:g17983153; GSPDB:GN00190  
A/Experimental source: strain 16M  
C/Genetics:  
A/Gene: BMEI1176  
A/Map position: 1

Query Match 2.5%; Score 6; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 FRASPP 187

Db 25 FRASPP 30

## RESULT 122

carbon storage regulator XP0125 [imported] - Xylella fastidiosa (strain 9a5C)  
C/Species: Xylella fastidiosa  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000  
C/Accession: D82844  
R/Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A/Reference number: AB2515; MUID:20385717; PMID:10910347  
A/Note: for a complete list of authors see reference number A59328 below  
A/Accession: D82844  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-76 <SIM>  
A/Cross-references: GB:AE003866; GB:AE003849; NID:g9104906; PIDN:AAF82938.1; GSPDB:GN001  
A/Experimental source: strain 9a5C  
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.; B  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir  
as-Neto, E.; Docena, C.; El-Dorfi, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins F  
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.F.  
Rodrigues, V.; Rosa, A.C.R.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsukako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

Query Match 2.5%; Score 6; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDSGS 11

Db 69 RGDSGS 74

## RESULT 123

T45103  
H<sup>+</sup>-transporting two-sector ATPase (EC 3.6.3.14) chain K [imported] - Methanosarcina mazei  
C:Species: Methanosarcina mazei  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 03-Jun-2002  
C:Accession: T45103  
R:Ruppert, C.; Wimmers, S.; Muller, V.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z22913  
A:Accession: T45103  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-80 <RUP>  
A:Cross-references: EMBL:U47274; PIDN:AAC06382.1  
A:Experimental source: strain G01  
C:Genetics:  
A:Gene: ahA  
C:Keywords: hydrolase

Query Match 2.5%; Score 6; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134

DB 22 AIAIAV 27

## RESULT 124

AF2255  
Hypothetical protein asl13597 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AF2255  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AF2255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB75296.1; PID:gl7132730; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asl13597

Query Match 2.5%; Score 6; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GELEKI 232

DB 35 GELEKI 40

## RESULT 125

AF1969  
Hypothetical protein asl1305 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AF1969  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AF1969  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-90 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB73262.1; PID:gl7130652; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asl1305

Query Match 2.5%; Score 6; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 TTAIPS 33

DB 82 TTAIPS 87

## RESULT 126

E83334  
Hypothetical protein PA2485 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83334  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: E83334  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-92 <STO>  
A:Cross-references: GB:AB004676; GB:AB004091; NID:G9948532; PIDN:AAG05873.1; GSPDB:GN00179  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2485

Query Match 2.5%; Score 6; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 FTSAGS 165

DB 28 FTSAGS 33

## RESULT 127

JH0716  
neuropeptide Y precursor - California sea hare  
C:Species: Aplysia californica (California sea hare)  
C>Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 21-Jul-2000  
C:Accession: JH0716  
R:Raybarr, S.M.; Garcia, P.D.; Roberts, R.; Eliassen, J.C.; Owens, D.F.; Maltby, D.; Myle. Neuron 9, 505-513, 1992  
A:Title: Identification and molecular cloning of a neuropeptide Y homolog that produces a hyperlocomotor response in the sea hare  
A:Reference number: JH0716; MUID:92398969; PMID:1524828  
A:Accession: JH0716  
A:Molecule type: mRNA  
A:Residues: 1-92 <RAJ>  
A:Cross-references: GB:M98854; NID:G155793; PIDN:AAA27772.1; PID:G155794  
A:Experimental source: abdominal ganglia  
C:Function:

A:Description: neuropeptide inducing a number of behavioral effects including stimulatory effects on feeding, swimming, and locomotion.  
C:Keywords: amidated carboxyl end; appetite; hormone; neuropeptide  
F11-21/Domain: signal sequence #status predicted <SIG>  
F12-61/Product: neuropeptide Y #status experimental <MAT>  
F12-92/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
F161/Modified site: amidated carboxyl end (phe) (amide in mature form from following glycosylation)

Query Match 2.5%; Score 6; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGDS 9



Db 62 GKEGDS 67  
|||||

## RESULT 128

T01876

hypothetical protein F8M12.13 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cross)

C/Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999

C/Accession: T01876

R/Madsen, C.; Graves, T.; Cotton, M.; Modde, T.

submitted to the EMBL Data Library, July 1998

A/Description: The sequence of A. thaliana F8M12.

A/Reference number: Z14450

A/Accession: T01876

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-93 &lt;MAD&gt;

A/Cross-references: EMBL:AF080118; NID:G3513725; PID:G3513729

A/Experimental source: cultivar Columbia

C/Genetics:

A/Map position: 4

A/Note: F8M12.13

## Query Match

Best Local Similarity 2.5%; Score 6; DB 2; Length 93;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233

Db 20 ELEKII 25

## RESULT 129

CCMP55

cytochrome c555 - Methylococcus capsulatus

C/Species: Methylococcus capsulatus

C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 03-Mar-2000

C/Accession: A23321; A05021

R/Amblar, R.P.; Dalton, H.; Meyer, T.E.; Bartsch, R.G.; Kamen, M.D.

Biochem. J. 233, 333-337, 1986

A/Title: The amino acid sequence of cytochrome c-555 from the methane-oxidizing bacterium

A/Reference number: A90328; MUID:86158741; PMID:3006666

A/Accession: A23321

A/Molecule type: protein

A/Residues: 1-96 &lt;AMB&gt;

A/Experimental source: strain Bath, NCIB 11132

C/Superfamily: cytochrome c6; cytochrome c6 homology

C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis

F:8-79/Domain: cytochrome c6 homology &lt;CVC&gt;

F:19-22/Binding site: heme (Cys) (covalent) #status experimental

F:723-59/Binding site: heme iron (His, Met) (axial ligands) #status predicted

## Query Match

Best Local Similarity 2.5%; Score 6; DB 1; Length 96;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AGELEK 231

Db 89 AGELEK 94

## RESULT 130

S30493

S1 protein - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999

C/Accession: S30493

R/Chestier, A.; Charnay, P.

DNA Seq. 2, 325-327, 1992

A/Title: Difference in the genomic organizations of the related transcription factors Sp

A/Reference number: S30493; MUID:92338398; PMID:1633330

A/Accession: S30493

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-101 &lt;CHE&gt;

A/Cross-references: EMBL:X60136; NID:G54158; PIDN:CAA42721.1; PID:e38120; PID:gl334268

## Query Match

Best Local Similarity 2.5%; Score 6; DB 2; Length 101;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AGSEGT 168

Db 48 AGSEGT 53

## RESULT 131

B90251

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C/Species: Sulfolobus solfataricus

C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C/Accession: B90251

R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.

attrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A/Description: Sulfolobus solfataricus complete genome.

A/Reference number: A99139

A/Accession: B90251

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-103 &lt;KUR&gt;

A/Cross-references: GB:AE006641; NID:gl3814179; PIDN:AAK41265.1; GSPDB:GN00155

C/Genetics:

A/Gene: SSO0994

## Query Match

Best Local Similarity 2.5%; Score 6; DB 2; Length 103;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233

Db 43 ELEKII 48

## RESULT 132

T44890

hypothetical protein MLCB22.15C [imported] - Mycobacterium leprae

C/Species: Mycobacterium leprae

C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 19-May-2000

C/Accession: T44890

R/Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1997

A/Reference number: Z22864

A/Accession: T44890

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-104 &lt;PAR&gt;

A/Cross-references: EMBL:Z98741; PIDN:CABL1380.1

A/Experimental source: cosmid B22

C/Genetics:

A/Note: MLCB22.15C

C/Superfamily: Mycobacterium leprae hypothetical protein MLCB22.15C

## Query Match

Best Local Similarity 2.5%; Score 6; DB 2; Length 104;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 ISRCTV 124

Db 79 ISRCTV 84

## RESULT 133

E72599

probable formylmethanofuran dehydrogenase APE1261 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: E72599  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1  
 A:Reference numbers: A72450; MUID:99310339; PMID:10382966  
 A:Accession: E72599  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-105 <KAW>  
 A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA080251.1; PID:G5104937  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE1261  
 C:Superfamily: Aeropyrum pernix probable formylmethanofuran dehydrogenase APE1261

Query Match 2.5%; Score 6; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPY 118  
 |||||  
 DB 12 RALEPY 17

RESULT 134  
 H86901  
 hypothetical protein ywjG [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: H86901  
 R:Boilotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis  
 A:Reference numbers: AB6625; MUID:21235186; PMID:111337471  
 A:Accession: H86901  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-106 <STO>  
 A:Cross-references: GB:AE005176; PID:G12725292; PIDN:AAK06314.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: ywjG

Query Match 2.5%; Score 6; DB 2; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKII 233  
 |||||  
 DB 46 ELEKII 51

RESULT 135  
 T42275  
 hypothetical protein - phage SPPI  
 C:Species: phage SPPI  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000  
 C:Accession: T42275  
 R:Alonso, J.C.; Luder, G.; Striege, A.C.; Chai, S.; Weise, F.; Trautner, T.A. Gene 204, 201-212, 1997  
 A:Title: The complete nucleotide sequence and functional organization of Bacillus subtilis phage SPPI  
 A:Reference numbers: 222137; MUID:98094274; PMID:9434185  
 A:Accession: T42275  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-107 <ALO>  
 A:Cross-references: EMBL:X97918; PIDN:CAA66585.1

Query Match 2.5%; Score 6; DB 2; Length 107;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 PLYSGF 45  
 |||||  
 DB 67 PLYSGF 72

RESULT 136  
 B71524  
 hypothetical protein CT357 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C:Species: Chlamydia trachomatis  
 C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
 C:Accession: B71524  
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis serotype D  
 A:Reference numbers: A71570; MUID:9900809; PMID:9784136  
 A:Accession: B71524  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-110 <ARN>  
 A:Cross-references: GB:AE001309; GB:AE001273; NID:G3328777; PIDN:AA067953.1; PID:G3328777  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: CT357

Query Match 2.5%; Score 6; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 LASPGS 177  
 |||||  
 DB 57 LASPGS 62

RESULT 137  
 T46071  
 hypothetical protein T18N14.130 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46071  
 R:Delveny, M.; Berger, C.; Cooke, R.; Greillet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.; submitted to the Protein Sequence Database, December 1999  
 A:Reference numbers: Z23013  
 A:Accession: T46071  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-110 <DEL>  
 A:Cross-references: EMBL:AL132968  
 A:Experimental source: cultivar Columbia; BAC clone T18N14  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 92/1  
 A>Note: T18N14.130

Query Match 2.5%; Score 6; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234  
 |||||  
 DB 83 LEKIIIS 88

RESULT 138  
 AF2540  
 hypothetical protein all7609 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120b;  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AF2540  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AF2540  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-112 <KUR>  
 A;Cross-references: GB:AP003602; PIDN:BAE77252.1; PID:gl7134694; GSPDB:GN00181  
 A;Experimental source: strain PCC 7120  
 C;Genetics:  
 A;Gene: all17609  
 A;Genome: plasmid

Query Match 2.5%; Score 6; DB 2; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 EEPFAS 185  
 |||||  
 DB 29 EEPFAS 34

RESULT 139  
 C31769  
 T-cell receptor delta-2 chain V region - human (fragment)  
 C;Species: Homo sapiens (man)  
 C;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 21-Jan-2000  
 C;Accession: C31769  
 R;Loh, E.Y.; Cwirla, S.; Serafini, A.T.; Phillips, J.H.; Lanier, L.L.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 9714-9718, 1988  
 A;Title: Human T-cell-receptor delta chain: genomic organization, diversity, and expression  
 A;Reference number: A94221; MUID:89071766; PMID:2974163  
 A;Accession: C31769  
 A;Molecule type: DNA  
 A;Residues: 1-113 <LOH>  
 A;Cross-references: GB:M23326; NID:G340877; PIDN:AAA61109.1; PID:G540457  
 C;Genetics:  
 A;Introns: 13/1  
 C;Superfamily: immunoglobulin V region; immunoglobulin homology  
 C;Keywords: T-cell receptor  
 F;33-113/Domain: immunoglobulin homology <IMM>

Query Match 2.5%; Score 6; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GFSELF 49  
 |||||  
 DB 6 GFSELF 11

RESULT 140  
 AH1784  
 hypothetical protein lin2822 [imported] - Listeria innocua (strain Clp11262)  
 C;Species: Listeria innocua  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AH1784  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, P.; Berche, P.; Bloecher, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makarewicz, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, O.; C. Schlueter, J.  
 A;Title: Comparative Genomics of Listeria species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AH1784  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-113 <GLA>  
 A;Cross-references: GB:AL592022; PIDN:CAC98048.1; PID:gl6415358; GSPDB:GN00178  
 A;Experimental source: strain Clp11262  
 C;Genetics:

A;Gene: lin2822

Query Match 2.5%; Score 6; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVPL 41  
 |||||  
 DB 90 EGTVPL 95

RESULT 141  
 B72648  
 hypothetical protein APE0616 - Aeropyrum pernix (strain K1)  
 C;Species: Aeropyrum pernix  
 C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C;Accession: B72648  
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yanazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Iwata, H.  
 DNA Res. 6, 83-101, 1999  
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
 A;Reference number: A72450; MUID:99310339; PMID:10382966  
 A;Accession: B72648  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-115 <KAW>  
 A;Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79596.1; PID:d1043372; PID:G5104188  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE0616

Query Match 2.5%; Score 6; DB 2; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASFG 176  
 |||||  
 DB 85 ALASFG 90

RESULT 142  
 T49363  
 hypothetical protein Bld1.170 [imported] - Neurospora crassa  
 C;Species: Neurospora crassa  
 C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
 C;Accession: T49363  
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, R.  
 submitted to the Protein Sequence Database, May 2000  
 A;Reference number: Z25022  
 A;Accession: T49363  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-119 <SCH>  
 A;Cross-references: EMBL:AL555927; GSPDB:GN00116; NCSP:Bld1.170  
 A;Experimental source: BAC clone Bld1; strain OR74A  
 C;Genetics:  
 A;Gene: NCSP:Bld1.170  
 A;Map position: 6  
 C;Superfamily: Neurospora crassa hypothetical protein Bld1.170

Query Match 2.5%; Score 6; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SCPEGT 38  
 |||||  
 DB 11 SCPEGT 16

RESULT 143  
 S10914  
 probable phosphoesterase (EC 3.1.1.-) - Synecococcus sp. (PCC 6301)  
 N;Alternate names: hypothetical protein 1 (16S rRNA 5' region)

C;Species: *Synechococcus* sp.  
 C;Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 24-Sep-1999  
 C;Accession: S10914  
 R;Kumano, M.; Tomioka, N.; Shinozaki, K.; Sugliura, M.  
 Mol. Gen. Genet. 202, 173-178, 1986  
 A;Title: Analysis of the promoter region in the rna operon from a blue-green alga, *Anacystis*  
 A;Reference number: S07311  
 A;Accession: S10914  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-119 <KUM>  
 A;Cross-references: EMBL:X03538; NID:G38918; PIDN:CAA7241.1; PID:G38921  
 C;Comment: This sequence has motifs characteristic of a variety of phosphoesterases.  
 C;Superfamily: unassigned probable phosphoesterases; phosphoesterase core homology  
 C;Keywords: hydrolase  
 F;4-82/Domain: phosphoesterase core homology <PEC>

Query Match 2.5%; Score 6; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133  
 |||||  
 Db 73 PAIAIA 78

RESULT 144  
 B95415  
 C;Species: *Sinorhizobium meliloti*  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C;Accession: B95415  
 R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A;Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*  
 A;Reference number: A95262; NUID:21396509; PMID:11481432  
 A;Accession: B95415  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-126 <KUR>  
 A;Cross-references: GB:AE006469; PIDN:AAK5884.1; PID:G14524393; GSPDB:GN00165  
 A;Experimental source: Strain 1021, megaplasmid pSymA  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Fuhrer, A.; Abola, P.; Ampe, P.; Barloy-Hubler,  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 Hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A;Reference number: A96039; NUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: SMA2277  
 A;Genome: plasmid

Query Match 2.5%; Score 6; DB 2; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SQTAT 31  
 |||||  
 Db 12 SQTAT 17

RESULT 145  
 A72388  
 C;Species: *Thermotoga maritima*  
 C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C;Accession: A72388  
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.  
 Nature 399, 323-329, 1999  
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se  
 A;Reference number: A72200; NUID:99287316; PMID:10360571  
 A;Accession: A72388  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-127 <ARN>  
 A;Cross-references: GB:AE001715; GB:AE000512; NID:G4980839; PIDN:AAD35426.1; PID:G49808  
 A;Experimental source: strain MSB8  
 C;Genetics:  
 A;Gene: TM0339  
 C;Superfamily: *Thermotoga maritima* hypothetical protein TM0339

Query Match 2.5%; Score 6; DB 2; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 FRASPF 187  
 |||||  
 Db 30 FRASPF 35

RESULT 146  
 S49637  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 19-Apr-2002  
 C;Accession: S49637  
 R;Gentles, S.; Bowman, S.  
 submitted to the EMBL Data Library, November 1994  
 A;Reference number: S49637  
 A;Accession: S49637  
 A;Molecule type: DNA  
 A;Residues: 1-128 <GEN>  
 A;Cross-references: EMBL:Z46660; NID:G575702; PID:G575713; GSPDB:GN00013; MIPS:YML090W  
 C;Genetics:  
 A;Gene: MIPS:YML090W  
 A;Cross-references: SGD:S0004555  
 A;Map position: 13L  
 C;Superfamily: *Saccharomyces cerevisiae* probable membrane protein YML090W  
 C;Keywords: transmembrane protein  
 F;6-22/Domain: transmembrane #status predicted <TM1>  
 F;55-71/Domain: transmembrane #status predicted <TM2>

Query Match 2.5%; Score 6; DB 2; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 FSPLEF 50  
 |||||  
 Db 4 FSPLEF 9

RESULT 147  
 A81062  
 C;Species: *Neisseria meningitidis*  
 C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C;Accession: A81062  
 R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.;  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.  
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vi  
 A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A;Reference number: A81000; NUID:20175755; PMID:10710307  
 A;Accession: A81062  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-134 <ET>  
 A;Cross-references: GB:AE002512; GB:AE002098; NID:G7226866; PIDN:AAF41971.1; PID:G72268  
 A;Experimental source: serogroup B, strain MC58

C;Genetics:  
A;Gene: NMB1619

Query Match 2.5%; Score 6; DB 2; Length 134;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DSGSPA 13  
|||||  
Db 31 DSGSPA 36

## RESULT 148

G81807  
hypothetical protein NMA1818 [imported] - Neisseria meningitidis (strain Z2491 serogroup  
C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C;Accession: G81807

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: G81807

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-134 <PAR>

A;Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85043.1; PID:g738045

A;Experimental source: serogroup A, strain Z2491

C;Genetics:

A;Gene: NMA1818

Query Match 2.5%; Score 6; DB 2; Length 134;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DSGSPA 13  
|||||  
Db 31 DSGSPA 36

## RESULT 149

C83908  
cytochrome aa3 quinol oxidase subunit IV qoxD [imported] - Bacillus halodurans (strain C  
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C;Accession: C83908

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: C83908

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-136 <STO>

A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05786.1; GSPDB:GN00

A;Experimental source: strain C-125

C;Genetics:

A;Gene: qoxD

Query Match 2.5%; Score 6; DB 2; Length 136;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAY 134  
|||||  
Db 85 AIAIAY 90

## RESULT 150

F70657  
hypothetical protein Rv2530c - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: F70657

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Comor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70657

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-139 <COL>

A;Cross-references: GB:Z83863; GB:AL123456; NID:g3261685; PIDN:CAB06178.1; PID:e290885;

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: Rv2530c

Query Match 2.5%; Score 6; DB 2; Length 139;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIA 133  
|||||  
Db 62 PAIAIA 67

Search completed: April 5, 2004, 07:39:26

Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:07:33 ; Search time 17 seconds  
(without alignments)  
747.360 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGSGSPATWTRGF.....KAGELEKIISRCQVCMKKRH 244

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	100.0	1670	1 CA34 HUMAN	Q01955 homo sapien
2	39	16.0	471	1 CA34 BOVIN	Q28084 bos taurus
3	17	7.0	754	1 CA54 CANFA	Q28247 canis famil
4	17	7.0	1669	1 CA14 HUMAN	P02462 homo sapien
5	17	7.0	1669	1 CA14 MOUSE	P02463 mus musculus
6	17	7.0	1685	1 CA34 HUMAN	P29400 homo sapien
7	11	4.5	1758	1 CA14 CAEEL	P17139 caenorhabdi
8	10	4.1	1712	1 CA24 HUMAN	P08572 homo sapien
9	9	3.7	1707	1 CA24 MOUSE	P08122 mus musculus
10	8	3.3	200	1 SODM AGABI	Q94456 agarius bi
11	8	3.3	281	1 SODF EACSU	Q35023 bacillus su
12	8	3.3	1691	1 CA54 HUMAN	Q14031 homo sapien
13	8	3.3	1775	1 CA14 DROME	P08120 drosophila
14	7	2.9	124	1 VB03 VACCC	P21000 vaccinia vi
15	7	2.9	137	1 RUVK OCEIH	Q89PT1 oceanobacil
16	7	2.9	151	1 RNB HSV2H	P89479 herpes simp
17	7	2.9	183	1 NUPM NEUCR	P21976 neurospora
18	7	2.9	201	1 SODM PROFR	P80293 propionibac
19	7	2.9	202	1 SODF METTM	Q60036 methanobact
20	7	2.9	210	1 SODF SULAC	Q08713 sulfobac
21	7	2.9	210	1 SODF SULSO	P80857 sulfobac
22	7	2.9	211	1 SODF ACTIAM	Q9P913 acidianus a
23	7	2.9	211	1 SODF PYRAE	Q93724 pyrobaculum
24	7	2.9	233	1 SODM YEAST	P00447 saccharomyc
25	7	2.9	245	1 SODM NEUCR	Q9Y783 neurospora
26	7	2.9	247	1 YACF SALTU	Q8XEX8 salmonella
27	7	2.9	259	1 CFAD MOUSE	P03953 mus musculus
28	7	2.9	370	1 HST4 YEAST	P53688 saccharomyc
29	7	2.9	409	1 YG45 YEAST	P50082 saccharomyc
30	7	2.9	441	1 YDM1 SCHPO	O13909 schizosacch
31	7	2.9	450	1 AROA MYCTU	P22487 mycobacteri
32	7	2.9	453	1 CA44 BOVIN	Q29442 bos taurus
33	7	2.9	488	1 GATE_RALSO	Q8V3C6 ralstonia s

34	7	2.9	550	1 HEMA IABAN	P03441 influenza a
35	7	2.9	550	1 HEMA_IADH1	P12582 influenza a
36	7	2.9	550	1 HEMA_IADH2	P12583 influenza a
37	7	2.9	550	1 HEMA_IADH3	P12584 influenza a
38	7	2.9	550	1 HEMA_IADH4	P12585 influenza a
39	7	2.9	550	1 HEMA_IADH5	P12586 influenza a
40	7	2.9	550	1 HEMA_IADH6	P12587 influenza a
41	7	2.9	550	1 HEMA_IADH7	P12588 influenza a
42	7	2.9	550	1 HEMA_IADH8	P12589 influenza a
43	7	2.9	550	1 HEMA_IADH9	P12590 influenza a
44	7	2.9	550	1 HEMA_IADH10	P12591 influenza a
45	7	2.9	550	1 HEMA_IADH11	P12592 influenza a
46	7	2.9	550	1 HEMA_IADH12	P12593 influenza a
47	7	2.9	550	1 HEMA_IADH13	P12594 influenza a
48	7	2.9	550	1 HEMA_IADH14	P12595 influenza a
49	7	2.9	550	1 HEMA_IADH15	P12596 influenza a
50	7	2.9	550	1 HEMA_IADH16	P12597 influenza a
51	7	2.9	550	1 HEMA_IADH17	P12598 influenza a
52	7	2.9	550	1 HEMA_IADH18	P12599 influenza a
53	7	2.9	550	1 HEMA_IADH19	P12600 influenza a
54	7	2.9	550	1 HEMA_IADH20	P12601 influenza a
55	7	2.9	550	1 HEMA_IADH21	P12602 influenza a
56	7	2.9	550	1 HEMA_IADH22	P12603 influenza a
57	7	2.9	550	1 HEMA_IADH23	P12604 influenza a
58	7	2.9	550	1 HEMA_IADH24	P12605 influenza a
59	7	2.9	550	1 HEMA_IADH25	P12606 influenza a
60	7	2.9	550	1 HEMA_IADH26	P12607 influenza a
61	7	2.9	550	1 HEMA_IADH27	P12608 influenza a
62	7	2.9	550	1 HEMA_IADH28	P12609 influenza a
63	7	2.9	550	1 HEMA_IADH29	P12610 influenza a
64	7	2.9	550	1 HEMA_IADH30	P12611 influenza a
65	7	2.9	550	1 HEMA_IADH31	P12612 influenza a
66	7	2.9	550	1 HEMA_IADH32	P12613 influenza a
67	7	2.9	550	1 HEMA_IADH33	P12614 influenza a
68	7	2.9	550	1 HEMA_IADH34	P12615 influenza a
69	7	2.9	550	1 HEMA_IADH35	P12616 influenza a
70	7	2.9	550	1 HEMA_IADH36	P12617 influenza a
71	7	2.9	550	1 HEMA_IADH37	P12618 influenza a
72	7	2.9	550	1 HEMA_IADH38	P12619 influenza a
73	7	2.9	550	1 HEMA_IADH39	P12620 influenza a
74	7	2.9	550	1 HEMA_IADH40	P12621 influenza a
75	7	2.9	550	1 HEMA_IADH41	P12622 influenza a
76	7	2.9	550	1 HEMA_IADH42	P12623 influenza a
77	7	2.9	550	1 HEMA_IADH43	P12624 influenza a
78	7	2.9	550	1 HEMA_IADH44	P12625 influenza a
79	7	2.9	550	1 HEMA_IADH45	P12626 influenza a
80	6	2.5	25	1 CHLY CARPA	P81241 carica papa
81	6	2.5	36	1 PSAL CHAGL	Q8M9X5 chaetospheae
82	6	2.5	46	1 LA89 LACAC	Q48501 lactobacill
83	6	2.5	65	1 PSB2 CYAPA	P17159 cyanophora
84	6	2.5	71	1 CSRA_XYLP	Q9PH21 xyella fas
85	6	2.5	96	1 CS55 METCA	P04369 methylococc
86	6	2.5	114	1 RSN MOUSE	Q99P87 mus musculu
87	6	2.5	116	1 CN4D MOUSE	Q01063 mus musculu
88	6	2.5	119	1 Y128 SYNPE	P05677 synechococ
89	6	2.5	128	1 YMJO YEAST	Q04501 saccharomyc
90	6	2.5	130	1 CAL2 MOUSE	Q99MP3 mus musculu
91	6	2.5	137	1 YAEU_PSEPU	P45588 pseudomonas
92	6	2.5	151	1 GLB8 CHITH	P02227 chironomus
93	6	2.5	152	1 TPC FATYE	P35622 patinopecte
94	6	2.5	158	1 RL2 ENTHI	O15574 entamoeba h
95	6	2.5	166	1 LIT4 HUMAN	P05451 homo sapien
96	6	2.5	166	1 LITB HUMAN	P48304 homo sapien
97	6	2.5	167	1 VB03 VACCV	Q01226 uareplasma
98	6	2.5	170	1 RECU_UREPA	Q9PJ4 ureaplasma
99	6	2.5	176	1 RL10 THEAC	Q9HJ33 thermoplasma
100	6	2.5	176	1 RL10 THEAC	P58299 thermoplasma
101	6	2.5	189	1 INAA BOVIN	P05007 bos taurus
102	6	2.5	199	1 SODN STAMP	Q99X82 staphylococ
103	6	2.5	200	1 SODM STRPN	Q59949 streptococ
104	6	2.5	202	1 CTDC HUMAN	Q9B11 homo sapien
105	6	2.5	202	1 SODM STRMU	P09738 streptococ
106	6	2.5	203	1 NU6M_TIRU	Q36836 trichophyto

107	6	2.5	206	1	SOD2_PLEBO	P50059	plectonema	180	6	2.5	350	1	YN14_YEAST	P53939	saccharomyc
108	6	2.5	207	1	JAG_EACHD	Q9rcac	bacillus ha	181	6	2.5	356	1	KPRA_HUMAN	Q14558	homo sapien
109	6	2.5	209	1	NGM_PARTE	P15600	paramecium	182	6	2.5	356	1	KPRA_MOUSE	Q9d0m1	mus musculus
110	6	2.5	210	1	GTSI_CABEL	Q09607	caenorhabdi	183	6	2.5	356	1	KPRA_RAT	Q63468	rattus norv
111	6	2.5	213	1	RS3_OCEIH	P59182	oceanobacil	184	6	2.5	361	1	PAX1_HUMAN	P15863	homo sapien
112	6	2.5	214	1	SODF_ARPE	Q9v8h8	aeropyrum p	185	6	2.5	369	1	DNAJ_HELPJ	Q9z1q2	helicobacte
113	6	2.5	217	1	FLA1_METJA	Q58301	methanococc	186	6	2.5	369	1	DNAJ_HELPY	O25890	helicobacte
114	6	2.5	217	1	FLA2_METJA	Q58302	methanococc	187	6	2.5	369	1	DNAJ_HELPY	O60356	homo sapien
115	6	2.5	217	1	RS3_EASTJ	P23309	bacillus st	188	6	2.5	369	1	KPRB_HUMAN	Q8r574	mus musculus
116	6	2.5	217	1	RS3_BACSU	P21465	bacillus su	189	6	2.5	369	1	KPRB_RAT	O08618	rattus norv
117	6	2.5	217	1	RS3_LACLA	Q9cdw8	lactococcus	190	6	2.5	371	1	CYB_LATCO	O08618	rattus norv
118	6	2.5	217	1	RS3_STAEM	Q99827	staphylococ	191	6	2.5	371	1	DUT_HSV11	P10234	herpes simp
119	6	2.5	217	1	RS3_STAEP	Q8crg6	staphylococ	192	6	2.5	373	1	CD62_METAC	O8t012	methanosaar
120	6	2.5	217	1	RS3_STRMU	P59186	streptococc	193	6	2.5	375	1	DNAJ_UREPA	O9pc82	ureaplasma
121	6	2.5	217	1	RS3_STRPN	Q9m437	streptococc	194	6	2.5	375	1	SR55_DROME	P26686	drosophila
122	6	2.5	217	1	RS3_STRPY	Q9a1w8	streptococc	195	6	2.5	379	1	DNAJ_LSGPN	P50025	legionella
123	6	2.5	218	1	RS3_LISMO	Q92713	listeria mo	196	6	2.5	388	1	OVAY_CHICK	P01014	gallus gall
124	6	2.5	219	1	RS3_BACAA	O81v54	bacillus an	197	6	2.5	388	1	YC09_KLEPN	Q48455	klebsiella
125	6	2.5	219	1	RS3_BACCR	O81j36	bacillus ce	198	6	2.5	390	1	TRPB_METTM	P26921	methanobact
126	6	2.5	219	1	RS3_BACHD	Q9z9k8	bacillus ha	199	6	2.5	392	1	TRB1_METTH	O27596	methanobact
127	6	2.5	222	1	SODM_HORSE	Q9x841	equus cabal	200	6	2.5	398	1	Y095_MYCGE	P47341	mycoplasma
128	6	2.5	223	1	PGC2_HUMAN	O15173	homo sapien								
129	6	2.5	223	1	THP2_RAT	Q8c5v7	rattus norv								
130	6	2.5	235	1	LIPB_MYCLE	O32961	mycobacteri								
131	6	2.5	248	1	SOD1_PLEBO	P50058	plectonema								
132	6	2.5	250	1	TRYP_PLEPL	P35034	pleuronecte								
133	6	2.5	260	1	GRAA_MOUSE	P11032	mus musculus								
134	6	2.5	262	1	ELNE_HUMAN	P12544	homo sapien								
135	6	2.5	267	1	ELNE_HUMAN	P08246	homo sapien								
136	6	2.5	268	1	ZUPT_OCEIH	Q8enq1	oceanobacil								
137	6	2.5	273	1	YDGD_ECOLI	P76176	escherichia								
138	6	2.5	281	1	UPK_STRPN	Q97sc8	streptococc								
139	6	2.5	282	1	AQP6_HUMAN	P11520	homo sapien								
140	6	2.5	287	1	TRUB_FUSNN	Q8r5x8	fusobacteri								
141	6	2.5	291	1	ZUPT_CAMJE	Q9pin2	campylobact								
142	6	2.5	292	1	EFTS_XANAC	Q8pmk6	xanthomonas								
143	6	2.5	292	1	EFTS_XANCP	Q8pav3	xanthomonas								
144	6	2.5	292	1	EFTS_XYLFA	Q9pad9	xyella fas								
145	6	2.5	292	1	EFTS_XYLFT	Q87a70	xyella fas								
146	6	2.5	298	1	VGLG_HRSV5	P27024	human respi								
147	6	2.5	299	1	FLR4_HUMAN	Q9g2z7	homo sapien								
148	6	2.5	301	1	CAPR_VERPE	P26950	versinia pe								
149	6	2.5	302	1	Y48_STAEP	Q8cte3	staphylococ								
150	6	2.5	303	1	Y367_RICPR	Q9z0g2	ricketsia								
151	6	2.5	308	1	HEM2_RHILO	Q98e17	rhizobium l								
152	6	2.5	311	1	FMT_BRAJA	Q89bp0	bradyrhizob								
153	6	2.5	313	1	M2OM_BOVIN	P22292	bos taurus								
154	6	2.5	313	1	M2OM_HUMAN	Q02978	homo sapien								
155	6	2.5	313	1	M2OM_MOUSE	Q9cr62	mus musculus								
156	6	2.5	313	1	M2OM_RAT	P97700	rattus norv								
157	6	2.5	313	1	MRW_PSESM	Q87wx7	pseudomonas								
158	6	2.5	313	1	MRW_PSEPK	Q8en84	pseudomonas								
159	6	2.5	316	1	LDH_STAEP	Q8cmz0	staphylococ								
160	6	2.5	325	1	Y893_MYCTU	Q10552	mycobacteri								
161	6	2.5	326	1	YJHS_ECOLI	P39370	escherichia								
162	6	2.5	328	1	P2Y6_HUMAN	Q15077	homo sapien								
163	6	2.5	329	1	HOLE_BACSU	P37540	bacillus su								
164	6	2.5	331	1	PME_ASPAC	Q12535	aspergillus								
165	6	2.5	332	1	OPPF_HAEIN	P45051	haemophilus								
166	6	2.5	334	1	OPPF_ECOLI	P77737	escherichia								
167	6	2.5	334	1	OPPF_SALTY	P08007	salmonella								
168	6	2.5	334	1	TRPD_STRPN	Q97p29	streptococc								
169	6	2.5	336	1	CMST_CRIGR	Q08520	cricketul								
170	6	2.5	336	1	CMST_MOUSE	Q61420	mus musculus								
171	6	2.5	337	1	CMST_HUMAN	P78382	homo sapien								
172	6	2.5	339	1	ANM1_SCHPO	Q9urx7	schizosacch								
173	6	2.5	344	1	ETFA_YEAST	Q12480	saccharomyc								
174	6	2.5	346	1	OMPA_ECOLI	P02934	escherichia								
175	6	2.5	346	1	YE97_METJA	Q58892	methanococc								
176	6	2.5	348	1	YGD9_YEAST	P53183	saccharomyc								
177	6	2.5	349	1	YCXF_PORPU	P51277	porphyra pu								
178	6	2.5	350	1	OMPA_ENTAE	P09146	enterobacte								
179	6	2.5	350	1	OMPA_SALTY	P02936	salmonella								

## ALIGNMENTS

## RESULT 1

## CA34\_HUMAN

ID CA34\_HUMAN STANDARD; PRT; 1670 AA.

AC Q01955; Q9BQT2;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).

## GN COL4A3

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

## RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

EX MEDLINE=94364994; PubMed=8083201;

RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;

RT "Complete primary structure of the human alpha 3(IV) collagen chain.

RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in

RT human tissues."

RL J. Biol. Chem. 269:23013-23017(1994).

## RN [2]

RP REVISIONS.

RA Leinonen A.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

## RN [3]

RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-1167;

RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; AND

RP CYS-1661; AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;

RP PRO-574; GLU-1269 AND PRO-1474.

RX MEDLINE=2064696; PubMed=1134255;

RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,

RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;

RT "Structure of the human type IV collagen gene COL4A3 and mutations in

RT autosomal Alport syndrome."

RL J. Am. Soc. Nephrol. 12:97-106(2001).

## RN [4]

RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=93015826; PubMed=1400291;

RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;

RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the

RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially

RT antigenic region at the triple helix/NC1 domain junction."

## RN [5]

RL J. Biol. Chem. 267:19780-19784(1992).

RP SEQUENCE OF 1453-1670 FROM N.A.





DR EMBL; AJ288528; CAC36101.1; JOINED.  
DR EMBL; AJ288529; CAC36101.1; JOINED.  
DR EMBL; AJ288530; CAC36101.1; JOINED.  
DR EMBL; AJ288531; CAC36101.1; JOINED.  
DR EMBL; AJ288532; CAC36101.1; JOINED.  
DR EMBL; AJ288533; CAC36101.1; JOINED.  
DR EMBL; AJ288534; CAC36101.1; JOINED.  
DR EMBL; AJ288535; CAC36101.1; JOINED.  
DR EMBL; AJ288536; CAC36101.1; JOINED.

Query Match 100.0%; Score 244; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 1.5e-242;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKXKRGDSGSPATWTRGFVTRISQTTAISCPEGVPLVYSGFSFLVQGNQRAHQD 60  
|||  
Db 1427 GLKXKRGDSGSPATWTRGFVTRISQTTAISCPEGVPLVYSGFSFLVQGNQRAHQD 1486  
|||

QY 61 LGTLGSLQRFMTPELFCNVNDVQNFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120  
|||  
Db 1487 LGTLGSLQRFMTPELFCNVNDVQNFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 1546  
|||

QY 121 RCTVCEGPAIAVAHSQTTDIPCPHGWLKKGFSFMFTSAGSECTGOALSPGSCLE 180  
|||  
Db 1547 RCTVCEGPAIAVAHSQTTDIPCPHGWLKKGFSFMFTSAGSECTGOALSPGSCLE 1606  
|||

QY 181 EFRASFLCHRGCTGCTNYNSYSFNLASINPERFRKPIPTVXAGELEKTIISRCQVCM 240  
|||  
Db 1607 EFRASFLCHRGCTGCTNYNSYSFNLASINPERFRKPIPTVXAGELEKTIISRCQVCM 1666  
|||

QY 241 KKRH 244  
|||  
Db 1667 KKRH 1670  
|||

## RESULT 2

CA34\_BOVIN STANDARD; PRT; 471 AA.  
ID CA34\_BOVIN  
AC Q28084;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Collagen alpha 3 (IV) chain (Fragment).  
GN COL4A3.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lens;  
RX MEDLINE=91093146; PubMed=1985905;  
RA Morrison K.E., Germino G.G., Readers S.T.;  
RT "Use of the polymerase chain reaction to clone and sequence a cDNA  
encoding the bovine alpha 3 chain of type IV collagen.";  
RL J. Biol. Chem. 266:34-39(1991).  
RN [2]  
RP SEQUENCE OF 227-258.  
RC TISSUE=Kidney;  
RX MEDLINE=90202779; PubMed=2318822;  
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;  
RT "Glomerular basement membrane. Identification of a fourth chain,  
alpha 4, of type IV collagen."  
RL J. Biol. Chem. 265:5466-5469(1990).  
RN [3]  
RP SEQUENCE OF 227-254.  
RX MEDLINE=8830844; PubMed=3417661;  
RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;  
RT "Identification of the Goodpasture antigen as the alpha 3 (IV) chain  
of collagen IV."  
RL J. Biol. Chem. 263:13374-13380(1988).  
RN [4]

RP SEQUENCE OF 227-244.  
RX MEDLINE=87222419; PubMed=2438283;  
RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,  
RA Hudson B.G.;  
RT "Localization of the Goodpasture epitope to a novel chain of basement  
membrane collagen.";  
RL J. Biol. Chem. 262:7874-7877(1987).  
CC -1- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.  
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1 (IV) -  
alpha 6 (IV), each of which can form a triple helix structure  
with 2 other chains to generate type IV collagen network.  
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the  
G-X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
CC -1- PM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -1- PM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M63139; AAA62708.1; -.  
DR PIR; A39024; A39024.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 4.  
DR ProDom; PD003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
FT NON\_TER 1 1  
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.  
FT SITE 239 471 NONHELICAL REGION (NC1).  
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).  
FT MOD\_RES 232 232 HYDROXYLATION.  
FT MOD\_RES 238 238 HYDROXYLATION.  
FT DISULFID 261 352 OR 349 (BY SIMILARITY).  
FT DISULFID 294 349 OR 352 (BY SIMILARITY).  
FT DISULFID 306 312 BY SIMILARITY.  
FT DISULFID 371 465 OR 463 (BY SIMILARITY).  
FT DISULFID 405 463 OR 466 (BY SIMILARITY).  
FT DISULFID 417 423 BY SIMILARITY.  
FT CONFLICT 253 253 S -> Y (IN REF. 3).  
SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;  
Query Match 16.0%; Score 39; DB 1; Length 471;  
Best Local Similarity 100.0%; Pred. No. 4e-32; Indels 0; Gaps 0;  
Matches 39; Conservative 0; Mismatches 0;  
QY 107 MAPITGRALEPYISRCTVCEGPAIAVAHSQTTDIPPCP 145  
|||  
Db 334 MAPITGRALEPYISRCTVCEGPAIAVAHSQTTDIPPCP 372  
|||

RESULT 3  
CA54\_CANFA STANDARD; PRT; 754 AA.  
ID CA54\_CANFA

Q28247;  
 01-NOV-1997 (Rel. 35, Created)  
 01-NOV-1997 (Rel. 35, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Collagen alpha 5(IV) chain (Fragment).  
 COL4A5.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Samoyed; TISSUE=Kidney;  
 RX MEDLINE=94224868; PubMed=8171024;  
 RA Zheng K., Thorner P.S., Marrano P., Bauml R., McInnes R.R.;  
 RT "Canine X chromosome-linked hereditary nephritis: a genetic model for  
 RT human X-linked hereditary nephritis resulting from a single base  
 RT mutation in the gene encoding the alpha 5 chain of collagen type  
 RT IV.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
 CC alpha 6(IV), each of which can form a triple helix structure with  
 CC 2 other chains to generate type IV collagen network.  
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
 CC X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -!- DISEASE: A defect in COL4A5 has been found to be the cause of  
 CC canine X-linked hereditary nephritis (HN), a disease similar to  
 CC that in humans (also referred to as Alport syndrome) characterized  
 CC by progressive renal failure and neurosensory deafness.  
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U07888; AAB60258.1; -;  
 DR PIR; A55267; A55267.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 8.  
 DR ProDom; PD000007; Clg\_helix; 1.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 FT NON\_TER 1 1  
 FT DOMAIN <1 530 TRIPLE-HELICAL REGION.  
 FT DOMAIN 531 >754 NONHELICAL REGION (NC1).  
 FT DISULFID 552 643 OR 640 (BY SIMILARITY).  
 FT DISULFID 585 640 OR 643 (BY SIMILARITY).  
 FT DISULFID 597 603 BY SIMILARITY.  
 FT DISULFID 662 ? OR 754 (BY SIMILARITY).  
 FT DISULFID 696 754 BY SIMILARITY.

FT DISULFID 708 714 BY SIMILARITY.  
 FT NON\_TER 754 754  
 SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;  
 Query Match 7.0%; Score 17; DB 1; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0;  
 Qy 84 VCNFASRNDYSYWLSTP 100  
 |||||  
 Db 602 VCNFASRNDYSYWLSTP 618  
 RESULT 4  
 CAL4 HUMAN  
 ID CAL4 HUMAN STANDARD; PRT; 1669 AA.  
 AC P02462;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89340433; PubMed=2701944;  
 RA Solininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;  
 RT "Structural organization of the gene for the alpha 1 chain of human  
 RT type IV collagen.";  
 RL J. Biol. Chem. 264:13565-13571(1989).  
 RN [2]  
 RP SEQUENCE OF 46-1257 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=8803584; PubMed=3691802;  
 RA Solininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
 RT "Complete primary structure of the alpha 1-chain of human basement  
 RT membrane (type IV) collagen.";  
 RL FEBS Lett. 225:188-194(1987).  
 RN [3]  
 RP SEQUENCE OF 1-943 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88029471; PubMed=3311751;  
 RA Brazel D., Oberhauser I., Dieringer H., Babel W., Glanville R.W.,  
 RA Deutzmann R., Kuehn K.;  
 RT "Completion of the amino acid sequence of the alpha 1 chain of human  
 RT basement membrane collagen (type IV) reveals 21 non-triplet  
 RT interruptions located within the collagenous domain.";  
 RL Eur. J. Biochem. 168:529-536(1987).  
 RN [4]  
 RP SEQUENCE OF 28-243  
 RX MEDLINE=8604708; PubMed=4043082;  
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
 RT region (7S domain) of the alpha 1 (IV) chain of human basement  
 RT membrane collagen.";  
 RL Eur. J. Biochem. 152:213-219(1985).  
 RN [5]  
 RP SEQUENCE OF 534-1447.  
 RX MEDLINE=85003629; PubMed=6434307;  
 RA Babel W., Glanville R.W.;  
 RT "Structure of human-basement-membrane (type IV) collagen. Complete  
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the  
 RT alpha 1(IV) chain.";  
 RL Eur. J. Biochem. 143:545-556(1984).  
 RN [6]  
 RP SEQUENCE OF 1256-1669 FROM N.A.  
 RX MEDLINE=85207819; PubMed=2581969;  
 RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;  
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV

RT procollagen reveal an unusual homology of amino acid sequences in two  
RT halves of the carboxyl-terminal domain.";  
RL J. Biol. Chem. 260:7681-7687(1985).  
RN [7]  
RP SEQUENCE OF 1259-1669 FROM N.A.  
RX MEDLINE=85216555; PubMed=2982422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RA Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
RT procollagen chains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89034231; PubMed=3182844;  
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
RT collagen are divergently encoded on opposite DNA strands and have an  
RT overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220(1988).  
RN [9]  
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RX MEDLINE=89005112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RT of basement-membrane type IV collagen.";  
RL Eur. J. Biochem. 176:617-624(1988).  
CC -1- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -1- PTM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -1- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -1- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M26576; AAA53098.1; JOINED.  
DR EMBL; J04217; AAA53098.1; JOINED.  
DR EMBL; M26550; AAA53098.1; JOINED.  
DR EMBL; M26540; AAA53098.1; JOINED.  
DR EMBL; M26542; AAA53098.1; JOINED.  
DR EMBL; M26543; AAA53098.1; JOINED.  
DR EMBL; M26544; AAA53098.1; JOINED.  
DR EMBL; M26545; AAA53098.1; JOINED.  
DR EMBL; M26546; AAA53098.1; JOINED.  
DR EMBL; M26547; AAA53098.1; JOINED.  
DR EMBL; M26537; AAA53098.1; JOINED.  
DR EMBL; M26538; AAA53098.1; JOINED.  
DR EMBL; M26548; AAA53098.1; JOINED.  
DR EMBL; M26549; AAA53098.1; JOINED.  
DR EMBL; M26551; AAA53098.1; JOINED.  
DR EMBL; M26552; AAA53098.1; JOINED.

DR EMBL; M26553; AAA53098.1; JOINED.  
DR EMBL; M26554; AAA53098.1; JOINED.  
DR EMBL; M26555; AAA53098.1; JOINED.  
DR EMBL; M26556; AAA53098.1; JOINED.  
DR EMBL; M26557; AAA53098.1; JOINED.  
DR EMBL; M26558; AAA53098.1; JOINED.  
DR EMBL; M26559; AAA53098.1; JOINED.  
DR EMBL; M26560; AAA53098.1; JOINED.  
DR EMBL; M26561; AAA53098.1; JOINED.  
DR EMBL; M26562; AAA53098.1; JOINED.  
DR EMBL; M26563; AAA53098.1; JOINED.  
DR EMBL; M26564; AAA53098.1; JOINED.  
DR EMBL; M26565; AAA53098.1; JOINED.  
DR EMBL; M26566; AAA53098.1; JOINED.  
DR EMBL; M26567; AAA53098.1; JOINED.  
DR EMBL; M26568; AAA53098.1; JOINED.  
DR EMBL; M26569; AAA53098.1; JOINED.  
DR EMBL; M26570; AAA53098.1; JOINED.  
DR EMBL; M26571; AAA53098.1; JOINED.  
DR EMBL; M26572; AAA53098.1; JOINED.  
DR EMBL; M26573; AAA53098.1; JOINED.  
DR EMBL; M26574; AAA53098.1; JOINED.  
DR EMBL; M26575; AAA53098.1; JOINED.  
DR EMBL; Y00706; CAA23075.1; -  
DR EMBL; X05561; CAA23075.1; -  
DR EMBL; M10940; AAA52006.1; -  
DR EMBL; M11315; AAA52042.1; -  
DR PR; S16876; CGHU4B.  
DR Genew; HGNC:2202; COL4A1.  
DR MM; 120130; -  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; Clg helix; 6.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL  
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .).  
FT DISULFID 1460 1551 OR 1548.  
FT DISULFID 1493 1548 OR 1551.  
FT DISULFID 1505 1511 OR 1662.  
FT DISULFID 1570 1665 OR 1662.  
FT DISULFID 1604 1662 OR 1665.  
FT DISULFID 1616 1622 SG -> KE (IN REF. 4).  
FT CONFLICT 237 238 Q -> K (IN REF. 4).  
FT CONFLICT 241 241 Q -> A (IN REF. 3).  
FT CONFLICT 319 319 N -> D (IN REF. 5).  
FT CONFLICT 719 719 D -> Y (IN REF. 5).  
FT CONFLICT 837 837 K -> P (IN REF. 5).  
FT CONFLICT 842 842 V -> W (IN REF. 2).  
FT CONFLICT 896 896 E -> Q (IN REF. 5).  
FT CONFLICT 904 904 S -> K (IN REF. 5).  
FT CONFLICT 914 914 S -> K (IN REF. 5).  
FT CONFLICT 998 998 K -> P (IN REF. 5).  
FT CONFLICT 1010 1010 S -> K (IN REF. 5).  
FT CONFLICT 1012 1012 K -> Q (IN REF. 5).  
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).  
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9B8A84 CRC64;

Query Match 7.0%; Score 17; DB 1; Length 1669;  
Best Local Similarity 100.0%; Pred. No. 5.1e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
DB 1510 VCNFASRNDYSYWLSTP 1526

RESULT 5  
CA14\_MOUSE  
ID CA14\_MOUSE STANDARD; PRT; 1669 AA.  
AC P02463;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Collagen alpha 1(IV) chain precursor.  
GN COL4A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RX MEDLINE=89197932; PubMed=2703490;  
RA Mathukumar G., Blumberg B., Kurkinen M.;  
RT "The complete primary structure for the alpha 1-chain of mouse  
collagen IV. Differential evolution of collagen IV domains.";  
RL J. Biol. Chem. 264:6310-6317(1989).  
RN [2]  
RX MEDLINE=89197932; PubMed=2703490;  
RA Mathukumar G., Blumberg B., Kurkinen M.;  
RT "The complete primary structure for the alpha 1-chain of mouse  
collagen IV. Differential evolution of collagen IV domains.";  
RL J. Biol. Chem. 264:6310-6317(1989).  
RN [3]  
RX MEDLINE=88112221; PubMed=3338568;  
RA Wood L., Theriault N., Vogeli G.;  
RT "cDNA clones completing the nucleotide and derived amino acid  
sequence of the alpha 1 chain of basement membrane (type IV) collagen  
from mouse.";  
RL FEBS Lett. 227:5-8(1988).  
RN [4]  
RX MEDLINE=86301886; PubMed=3755692;  
RA Nath P., Laurent M., Horn E., Sobel M.B., Zon G., Vogeli G.;  
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a  
synthetic oligodeoxynucleotide.";  
RL Gene 43:301-304(1986).  
RN [5]  
RX MEDLINE=85127033; PubMed=2578961;  
RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;  
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of  
the alpha 1(IV) chain of basement membrane collagen as derived from  
complementary DNA.";  
RL Eur. J. Biochem. 147:217-224(1985).  
RN [6]  
RX MEDLINE=87250460; PubMed=3597383;  
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
Saus J., Pihlajaniemi T.;  
RT "Extensive homology between the carboxyl-terminal peptides of mouse  
alpha 1(IV) and alpha 2(IV) collagen.";  
RL J. Biol. Chem. 262:8496-8499(1987).  
RN [7]  
RX MEDLINE=86196099; PubMed=3009468;  
RA Sakurai Y., Sullivan M., Yamada Y.;  
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
collagen genes.";  
RL J. Biol. Chem. 261:6654-6657(1986).  
RN [8]  
RX MEDLINE=89066738; PubMed=3198626;  
RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
RT "Head-to-head arrangement of murine type IV collagen genes.";  
RL J. Biol. Chem. 263:19274-19277(1988).  
RN [9]  
RX MEDLINE=89071759; PubMed=3200851;  
RA Burbello P.D., Martin G.R., Yamada Y.;  
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
bidirectional promoter and a shared enhancer.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).  
RN [9]  
RX MEDLINE=88243724; PubMed=3379041;  
RA Killen P.D., Burbello P., Sakurai Y., Yamada Y.;  
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
collagen chain and the corresponding region of the gene.";  
RL J. Biol. Chem. 263:8706-8709(1988).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
glomerular basement membranes (GBM), forming a 'chicken-wire'  
meshwork together with laminins, proteoglycans and entactin/  
nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
alpha 6(IV), each of which can form a triple helix structure with  
2 other chains to generate type IV collagen network.  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
domain (NC1) at their C-terminus, frequent interruptions of the G-  
X-Y repeats in the long central triple-helical domain (which may  
cause flexibility in the triple helix), and a short N-terminal  
triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
are involved in inter- and intramolecular disulfide bonding. 12 of  
these, located in the NC1 domain, are conserved in all known type  
IV collagens.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
DR EMBL; J03758; AAA37439.1; -  
DR EMBL; M23333; AAA51625.1; -  
DR EMBL; J04694; AAA50292.1; -  
DR EMBL; X06777; CAA29946.1; -  
DR EMBL; X02201; CAA26132.1; -  
DR EMBL; M15832; AAA37340.1; -  
DR EMBL; M14042; AAA37342.1; -  
DR EMBL; M12879; AAA37343.1; -  
DR EMBL; M13024; -; NOT ANNOTATED CDS.  
DR EMBL; M13025; -; NOT ANNOTATED CDS.  
DR EMBL; M13026; AAA37344.1; -  
DR EMBL; M13027; AAA37345.1; -  
DR EMBL; M13043; AAA37346.1; -  
DR EMBL; J04448; AAA37437.1; -  
DR PIR; A33525; CGMS4B.  
DR MGI; 88454; Col4a1.  
DR GO; GO:0005604; C:basement membrane; IDA.  
DR InterPro; IPR008161; C1q helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 23.  
DR ProDom; PD000007; C1q helix; 6.  
DR ProDom; PD003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 27  
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).

```
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 V -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E5205859 CRC64;

Query Match 7.0%; Score 17; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 5,1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
Db 1510 VCNFASRNDYSYWLSTP 1526

RESULT 6
CAS4_HUMAN STANDARD; PRT; 1685 AA.
ID P29400; Q16006; Q16126;
AC 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5 (IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
[2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=92318923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
[3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjola-Erila E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5 (IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
[4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
[5]
```

```
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoesly M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
[6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola-Erila E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5 (IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
[7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NCI domain.";
RL Kidney Int. 44:1316-1321(1993).
[9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
[10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5 (IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
[11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: Characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
[12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
[13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]
```

RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;  
RP ARG-869; ARG-872 AND CYS-1241.  
RX MEDLINE=95322976; PubMed=7599631;  
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
RT "Detection of 12 novel mutations in the collagenous domain of the  
RT COL4A5 gene in Alport syndrome patients.";  
RL Hum. Mutat. 5:197-204(1995).  
RN [15]  
RP VARIANTS AS ARG-1649.  
RX MEDLINE=96213750; PubMed=8651292;  
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
RA Denison J.C., Fain P.R., Gregory M.C.;  
RT "A mutation causing Alport syndrome with tardive hearing loss is  
RT common in the western United States.";  
RL Am. J. Hum. Genet. 58:1157-1165(1996).  
RN [16]  
RP VARIANTS AS.  
RX MEDLINE=96213754; PubMed=8651296;  
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
RA Savi M., Ballabio A., de Marchi M.;  
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
RT exons of the COL4A5 Gene.";  
RL Am. J. Hum. Genet. 58:1192-1204(1996).  
RN [17]  
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
RP MET-1428.  
RX MEDLINE=97094179; PubMed=8940267;  
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,  
RA Glatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
RA Gubler M.-C., Antignac C.;  
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
RT syndrome.";  
RL Am. J. Hum. Genet. 59:1221-1232(1996).  
RN [18]  
RP VARIANTS AS ASP-1498.  
RX MEDLINE=96233932; PubMed=8829632;  
RA Tverskaya S., Bobrynya V., Tsalykova F., Ignatova M.,  
RA Krasnopolskaya X., Evgrafov O.;  
RT "Substitution of A1498D in noncollagen domain of  $\alpha 5$ (IV) collagen  
RT chain associated with adult-onset X-linked Alport syndrome.";  
RL Hum. Mutat. 7:149-150(1996).  
RN [19]  
RP VARIANTS AS GLN-1677.  
RX MEDLINE=97295089; PubMed=9150741;  
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
RT "Common ancestry of three Ashkenazi-American families with Alport  
RT syndrome and COL4A5 R1677Q";  
RL Hum. Genet. 99:681-684(1997).  
RN [20]  
RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1517  
RP AND ASP-1596.  
RX MEDLINE=98112435; PubMed=9452056;  
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;  
RT "Missense mutations in the COL4A5 gene in patients with X-linked  
RT Alport syndrome.";  
RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
RN [21]  
RP VARIANTS AS VAL-420; 456-PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
RP 802-GLY-PRO-807 DEL; ARG-889; CYS-941; SER-1030; SER-1066; ASP-1143;  
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.  
RX MEDLINE=99063529; PubMed=9848783;  
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumeilius T., Hertz J.M.,  
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,  
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;  
RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
RT Alport syndrome using PCR and direct DNA sequencing.";  
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
RN [22]  
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;

RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
RX MEDLINE=20030197; PubMed=10561141;  
RA Inoue Y., Nishio H., Shirakawa T., Nakarishi K., Nakamura H.,  
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
RT "Detection of mutations in the COL4A5 gene in over 90% of male  
RT patients with X-linked Alport's syndrome by RT-PCR and direct  
RT sequencing.";  
RL Am. J. Kidney Dis. 34:854-862(1999).  
RN [23]  
RP VARIANT AS ARG-822.  
Query Match 7.0%; Score 17; DB 1; Length 1685;  
Best Local Similarity 100.0%; Pred. No. 5.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 84 VCNFASRNDYSYWLSTP 100  
DB 1526 VCNFASRNDYSYWLSTP 1542  
CA14\_CABEL STANDARD; PRT; 1758 AA.  
AC P17139;  
DC 01-AUG-1990 (Rel. 15, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(IV) chain precursor.  
GN EMB-9 OR CLB-2 OR K04H4.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Bristol N2;  
RC MEDLINE=91141582; PubMed=1996137;  
RA Guo X., Johnson J.J., Kramer J.M.;  
RT "Embryonic lethality caused by mutations in basement membrane  
RT collagen of C. elegans.";  
RL Nature 349:707-709(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Bristol N2;  
RC MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Staden R.,  
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden K.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Woldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [3]  
RP REVISIONS.  
RA Durbin R.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1446-1758 FROM N.A.  
RX STRAIN=Bristol N2;  
RC MEDLINE=90080929; PubMed=2793871;  
RA Guo X., Kramer J.M.;  
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen  
RT genes are located on separate chromosomes.";  
RL J. Biol. Chem. 264:17574-17582(1989).  
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.  
CC -1- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.



Type IV collagen forms a mesh-like network linked through intermolecular interactions between 7S domains and between NC1 domains.

-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

-!- PPM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

-!- PPM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

-!- DISEASE: Mutations in this gene cause temperature-sensitive lethality during late embryogenesis.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X56979; CAA40299.1; -;  
EMBL; Z27078; CAA81584.3; -;  
EMBL; J05067; AAB59179.1; -;  
PIR; S40991; S40991;  
WormPep; K04H4.1; CB32462.  
InterPro; IPR008161; C1g.Helix.  
InterPro; IPR008160; Collagen.  
InterPro; IPR001442; Procollagen4\_C.  
Pfam; PF01413; C4; 2.  
Pfam; PF01391; Collagen; 27.  
ProDom; PD000007; C1g.Helix; 11.  
ProDom; PD003923; Procollagen4; 1.  
SMART; SM00111; C4; 2.  
Extracellular matrix; Connective tissue; Basement membrane;  
Repeat; Hydroxylation; Glycoprotein; Signal.  
SIGNAL 1 20  
PROPEP 21 2194 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
CHAIN 2195 1758 COLLAGEN ALPHA 1(IV) CHAIN.  
DOMAIN 195 1529 TRIPLE-HELICAL REGION.  
DOMAIN 1530 1758 NONHELICAL REGION (NC1).  
DISULFID 1549 1640 OR 1637 (BY SIMILARITY).  
DISULFID 1582 1637 OR 1640 (BY SIMILARITY).  
DISULFID 1594 1600 BY SIMILARITY.  
DISULFID 1659 1754 OR 1751 (BY SIMILARITY).  
DISULFID 1693 1751 OR 1754 (BY SIMILARITY).  
DISULFID 1705 1711 BY SIMILARITY.  
VARIANT 402 402 G -> E (IN MUTANT G34).  
VARIANT 408 408 G -> E (IN MUTANT G23/HG70).  
CONFLICT 302 391 LDNGKRGKGVGNGYGEKSGQGLGTPGYPTKGGAGE  
RGVPRGFGSDGCGPEGLGEGTGEAGPHCAQGFQVQSGK  
GLPGRDGL -> AGQVSIQPNKKLFLFCRVNVERTEQSE  
IMKRDPEKNKDLLEQDTQLREGLNQDTQEDQVSKETV  
DRKHLEKLEKVRDLDMELKDSFTFEKAKQCDQVMVVS (IN  
REF. 2).  
G -> R (IN REF. 2).  
P -> R (IN REF. 2).  
D -> V (IN REF. 2).  
P -> Q (IN REF. 4).  
P -> L (IN REF. 2).  
CONFLICT 581 581  
CONFLICT 768 768  
CONFLICT 830 830  
CONFLICT 1514 1514  
CONFLICT 1722 1722  
SEQUENCE 1758 AA; 170857 MW; 7083D9AF63E05D45 CRC64;  
Query Match 4.5%; Score 11; DB 1; Length 1758;  
Best Local Similarity 100.0%; Pred. No. 0.008;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 174 SPGSCLEEFRA 184  
Db 1689 SPGSCLEEFRA 1699

RESULT 8  
CA24 HUMAN  
ID CA24 HUMAN STANDARD; PRT; 1712 AA.  
AC P08572;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Collagen alpha 2(IV) chain precursor.  
GN COL4A2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89066769; PubMed=3198637;  
RA Hostikka S.L., Tryggvason K.;  
RT "The complete primary structure of the alpha 2 chain of human type IV  
collagen and comparison with the alpha 1(IV) chain.";  
RL J. Biol. Chem. 263:19488-19493(1988).  
RN [2]  
RP SEQUENCE OF 1-1042 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88151998; PubMed=3345760;  
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;  
RT "Human basement membrane collagen (type IV). The amino acid sequence  
of the alpha 2(IV) chain and its comparison with the alpha 1(IV)  
chain reveals deletions in the alpha 1(IV) chain.";  
RL Eur. J. Biochem. 172:35-42(1988).  
RN [3]  
RP SEQUENCE OF 1254-1712 FROM N.A.  
RX MEDLINE=87219158; PubMed=3582677;  
RA Hostikka S.L., Kurkinen M., Tryggvason K.;  
RT "Nucleotide sequence coding for the human type IV collagen alpha 2  
chain cDNA reveals extensive homology with the NC-1 domain of alpha 1  
(IV) but not with the collagenous domain or 3'-untranslated region.";  
RL FEBS Lett. 216:281-286(1987).  
RN [4]  
RP SEQUENCE OF 1451-1485 FROM N.A.  
RX MEDLINE=87092438; PubMed=3025878;  
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;  
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and  
alpha 2 (IV) chains map to the distal long arm of chromosome 13.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).  
RN [5]  
RP SEQUENCE OF 1486-1712 FROM N.A.  
RX MEDLINE=87250571; PubMed=2439508;  
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;  
RT "Duplication of type IV collagen COOH-terminal repeats and species-  
specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";  
RL J. Biol. Chem. 262:9231-9238(1987).  
RN [6]  
RP SEQUENCE OF 1-33 FROM N.A.  
RX MEDLINE=89034231; PubMed=3182844;  
RA Sohininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
collagen are divergently encoded on opposite DNA strands and have an  
overlapping promoter region.";  
RL J. Biol. Chem. 263:17217-17220(1988).  
RN [7]  
RP SEQUENCE OF 1-33 FROM N.A.  
RX MEDLINE=89030632; PubMed=2846280;  
RA Poeschl E., Pollner R., Kuehn K.;  
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human  
basement membrane collagen type IV are arranged head-to-head and  
separated by a bidirectional promoter of unique structure.";  
RL EMBO J. 7:2687-2695(1988).  
RN [8]  
RP SEQUENCE OF 1-33 FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=93305049; PubMed=8317999;  
RN

RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;  
 RT "Identification of a novel sequence element in the common promoter  
 RT region of human collagen type IV genes, involved in the regulation of  
 RT divergent transcription.";  
 RL Biochem. J. 292:687-695(1993).  
 RN [9]  
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.  
 RC TISSUE=Placenta;  
 RX MEDLINE=89005112; PubMed=2844531;  
 RA Siebold B., Deutzmann R., Kuehn K.;  
 RT "The arrangement of intra- and intermolecular disulfide bonds in the  
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
 RT of basement-membrane type IV collagen.";  
 RL Eur. J. Biochem. 176:617-624(1988).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC meshwork together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
 CC alpha 6(IV), each of which can form a triple helix structure  
 CC with 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the  
 CC G-X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05562; CAA29076.1; -;  
 DR EMBL; X05610; CAA29098.1; -;  
 DR EMBL; J02760; AAA58422.1; -;  
 DR EMBL; M36963; AAA53059.1; -;  
 DR EMBL; X12784; CAA31275.1; -;  
 DR EMBL; J04217; AAA53097.1; -;  
 DR PIR; A32024; CGH02B.  
 DR Genew; HGNC:2203; COL4A2.  
 DR MIM; 120030; -;  
 DR GO; GO:0005587; C:collagen type IV; TAS.  
 DR GO; GO:0005201; P:extracellular matrix structural constituent; TAS.  
 DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.  
 DR InterPro; IPR008161; C1g helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 24.  
 DR ProDom; PD000007; C1g\_helix; 7.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL  
 FT PROPEP 1 25  
 FT CHAIN 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT DOMAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).  
 FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).  
 FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).  
 FT DISULFID 1549 1555 BY SIMILARITY.  
 FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).  
 FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).

FT DISULFID 1558 1665 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).  
 R -> P (IN REF. 2).  
 FT CONFLICT 471 471 A -> G (IN REF. 2).  
 FT CONFLICT 683 683 M -> I (IN REF. 5).  
 FT CONFLICT 1575 1575 G -> H (IN REF. 9).  
 FT CONFLICT 1663 1663 H -> G (IN REF. 9).  
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
 SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
 Query Match 4.1%; Score 10; DB 1; Length 1712;  
 Best Local Similarity 100.0%; Pred. No. 0.083;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 PAIAIAVHSQ 137  
 Db 1596 PAIAIAVHSQ 1605  
 RESULT 9  
 CA24 MOUSE  
 ID CA24 MOUSE STANDARD; PRT; 1707 AA.  
 AC P08122; Q61375;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 GN COL4A2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89197933; PubMed=2703491;  
 RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumaran G.,  
 RA Pihlajaniemi T., Kurkinen M.;  
 RT "The complete primary structure of mouse alpha 2(IV) collagen.  
 RT Alignment with mouse alpha 1(IV) collagen.";  
 RL J. Biol. Chem. 264:6318-6324(1989).  
 RN [2]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89066738; PubMed=3198626;  
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
 RT "Head-to-head arrangement of murine type IV collagen genes.";  
 RL J. Biol. Chem. 263:19274-19277(1988).  
 RN [3]  
 RP SEQUENCE OF 970-1480 FROM N.A.  
 RX MEDLINE=86220192; PubMed=3011432;  
 RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,  
 RA Deutzmann R., Timpl R., Kuehn K.;  
 RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-  
 RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)  
 RT chain and its comparison with the alpha 1(IV) chain.";  
 RL Eur. J. Biochem. 157:49-56(1986).  
 RN [4]  
 RP SEQUENCE OF 1480-1707 FROM N.A.  
 RX MEDLINE=87054581; PubMed=3780963;  
 RA Schwarz-Magdolen U., Oberbauer I., Kuehn K.;  
 RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of  
 RT collagen IV and its comparison with alpha 1(IV).";  
 RL FEBS Lett. 208:203-207(1986).  
 RN [5]  
 RP SEQUENCE OF 1481-1707 FROM N.A.  
 RX MEDLINE=87250460; PubMed=3597383;  
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
 RA Saus J., Pihlajaniemi T.;  
 RT "Extensive homology between the carboxyl-terminal peptides of mouse  
 RT alpha 1(IV) and alpha 2(IV) collagen.";  
 RL J. Biol. Chem. 262:8496-8499(1987).  
 RN [6]  
 RP SEQUENCE OF 1041-1489 FROM N.A.  
 RX MEDLINE=87005245; PubMed=3758345;  
 RA Vogeli G., Horn E., Carter J., Kaytes P.S.;

"Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";

FEBS Lett. 206:29-32(1986).

[7]

SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.

RX MEDLINE=85296379; PubMed=3839908;

RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;

RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";

RL Nature 317:177-179(1985).

[8]

SEQUENCE OF 1-60 FROM N.A.

RX MEDLINE=89071759; PubMed=3200851;

RA Buelo P.D., Martin G.R., Yamada Y.;

RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).

CC -!- FUNCTION: Type IV collagen is the major structural component of glomerular basement membranes (GBM), forming a 'chicken-wire' meshwork together with laminins, proteoglycans and entactin/nidogen.

CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) - alpha 6(IV), each of which can form a triple helix structure with 2 other chains to generate type IV collagen network.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M23334; AAA51626.1; -

DR EMBL; M23333; AAA51626.1; JOINED.

DR EMBL; J04695; AAA50293.1; -

DR EMBL; J04448; AAA37438.1; -

DR EMBL; X04647; CAA28308.1; -

DR EMBL; M15833; AAA37341.1; -

DR EMBL; X04410; CAA27998.1; -

DR EMBL; X02896; CAA26655.1; -

DR EMBL; X02897; CAB51614.1; -

DR EMBL; X02898; CAA26657.1; -

DR EMBL; X02899; CAA26658.1; -

DR PIR; A33526; A33526.

DR MGD; MGI:88455; Col4a2.

GO; GO:0005604; C:basement membrane; IDA.

DR InterPro; IPR008161; C1g helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR001442; Procollagen4\_C.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 21.

DR ProDom; PD000007; C1g helix; 7.

DR ProDom; PD003923; ProcollagenC4; 1.

DR SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Signal.

FT SIGNAL 1 25

FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 184 1707 COLLAGEN ALPHA 2(IV) CHAIN.

FT DOMAIN 184 1479 TRIPLE-HELICAL REGION.

FT DOMAIN 1480 1707 NONHELICAL REGION (NC1).

FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).

FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).

FT DISULFID 1544 1585 BY SIMILARITY.

FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).

FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).

FT DISULFID 1653 1660 BY SIMILARITY.

FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1370 1370 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 1051 1051 P -> R (IN REF. 6).

FT CONFLICT 1097 1097 S -> G (IN REF. 7).

FT CONFLICT 1171 1171 G -> S (IN REF. 6).

FT CONFLICT 1179 1179 P -> R (IN REF. 6).

FT CONFLICT 1241 1241 Q -> E (IN REF. 6).

FT CONFLICT 1328 1328 P -> A (IN REF. 6).

FT CONFLICT 1573 1573 V -> L (IN REF. 4).

FT CONFLICT 1623 1623 Y -> H (IN REF. 4).

SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query Match 3.7%; Score 9; DB 1; Length 1707;

Best Local Similarity 100.0%; Pred. No. 0.88;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCMK 241

Db 1697 ISRCQVCMK 1705

RESULT 10

SODM AGABI

ID SODM AGABI STANDARD; PRT; 200 AA.

AC Q2P4T6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Superoxide dismutase [Mn], mitochondrial precursor (RC 1.15.1.1).

GN SOD.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI\_TaxID=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Hortet U3;

RA Eastwood D.C., Bains N.K., Henderson J., Burton K.S.;

RT "Oxidative stress in the harvested mushroom, *Agaricus bisporus*.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase family.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AJ404469; CAB94731.1; -

DR HSSP; P04179; 1VAR.

DR InterPro; IPR001189; SODismutase.

DR Pfam; PF00081; sodfe; 1.

DR Pfam; PF02777; sodfe; C; 1.

DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom; PD000475; SODismutase; 2.

DR PROSITE; PS00088; SOD MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Mitochondrion;

KW Transit peptide.

FT TRANSIT 1 1 ? MITOCHONDRION (POTENTIAL).

```

FT CHAIN ? 200 SUPEROXIDE DISMUTASE [MN]
FT METAL 27 27 MANGANESE (BY SIMILARITY)
FT METAL 72 72 MANGANESE (BY SIMILARITY)
FT METAL 157 157 MANGANESE (BY SIMILARITY)
FT METAL 161 161 MANGANESE (BY SIMILARITY)
SQ SEQUENCE 200 AA; 22194 MW; 975981DD1F674F19 CRC64;

Query Match 3.3%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYISR 121
DB 14 ALEPYISR 21

RESULT 11
SODF_BACSU STANDARD; PRT; 281 AA.
AC Q35023;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable superoxide dismutase [rel (EC 1.15.1.1)].
GN SODF OR BSU9330.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RL the terC and oshAB loci cloned in a yeast artificial chromosome.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrarri E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medevic C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Neone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle D., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 iron ion per subunit (By similarity).

```

```

CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF027868; AAB84442.1;
CC DR EMBL; Z99114; CAB13825.1;
CC DR PIR; C69709; C69709.
CC DR HSSP; P80293; 1A0M.
CC DR Subtilist; B312876; sodF.
CC DR InterPro; IPR001189; SODismutase.
CC DR Pfam; PF00081; sodfe; 1.
CC DR Pfam; PF02777; sodfe C; 1.
CC DR PRINTS; PR01703; MNSODISMUTASE.
CC DR ProDom; PD000475; SODismutase; 1.
CC DR PROSITE; PS00088; SOD_MN; 1.
CC KW Oxidoreductase; Metal-binding; Iron; Complete proteome.
FT METAL 104 104 IRON (BY SIMILARITY).
FT METAL 152 152 IRON (BY SIMILARITY).
FT METAL 236 236 IRON (BY SIMILARITY).
FT METAL 240 240 IRON (BY SIMILARITY).
SQ SEQUENCE 281 AA; 33477 MW; 7F36AC0A60E74DB0 CRC64;

Query Match 3.3%; Score 8; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYISR 121
DB 91 ALEPYISR 98

RESULT 12
CA64_HUMAN STANDARD; PRT; 1691 AA.
ID CA64_HUMAN STANDARD; PRT; 1691 AA.
AC Q1403; Q12823; Q14053; Q9NOM5; Q9NTX3; Q9UJ76; Q9UMG6; Q9V4L4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Eyes, and Kidney.
RX MEDLINE=94171779; PubMed=8125972;
RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RN J. Biol. Chem. 269:7520-7526(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=94230418; PubMed=8175748;
RA Zhou J., Ding M., Zhao Z., Reiders S.T.;
RT "Complete primary structure of the sixth chain of human basement
RT membrane collagen, alpha 6(IV). Isolation of the cDNAs for alpha 6(IV)
RT and comparison with five other type IV collagen chains.";
RN J. Biol. Chem. 269:13193-13199(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS ALA-455 AND
RP LYS-1110.
RX MEDLINE=96299642; PubMed=8661006;
RA Zhang X., Zhou J., Reiders S.T., Tryggvason K.;

```

RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
RT in Alport syndrome-associated leiomyomatosis.";  
RL Genomics 33:473-479(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Bird C., Grafham D., Lawlor S., Wilson S.,  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-542 FROM N.A. (ISOFORM A).  
RX MEDLINE=9361972; PubMed=8356449;  
RA Zhou J., Mochizuki T., Smeets H., Antignac C., Laurila P.,  
RL de Paape A., Tryggvason K., Reiders S.T.;  
RT "Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in  
RL inherited smooth muscle tumors.";  
RL Science 261:1167-1169(1993).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=Q14031-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q14031-2; Sequence=VSP\_001174;  
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NCL) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- PM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NCL domain, are conserved in all known type  
CC IV collagens.  
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D21337; BAA04809.1; -;  
DR EMBL; U04845; AAA19569.2; -;  
DR EMBL; U47004; AAB19038.1; -;  
DR EMBL; U46959; AAB19038.1; JOINED.  
DR EMBL; U46961; AAB19038.1; JOINED.  
DR EMBL; U46962; AAB19038.1; JOINED.  
DR EMBL; U46963; AAB19039.1; JOINED.  
DR EMBL; U46964; AAB19038.1; JOINED.  
DR EMBL; U46965; AAB19038.1; JOINED.  
DR EMBL; U46966; AAB19038.1; JOINED.  
DR EMBL; U46967; AAB19038.1; JOINED.  
DR EMBL; U46968; AAB19038.1; JOINED.  
DR EMBL; U46969; AAB19038.1; JOINED.  
DR EMBL; U46970; AAB19038.1; JOINED.  
DR EMBL; U46971; AAB19038.1; JOINED.  
DR EMBL; U46972; AAB19038.1; JOINED.  
DR EMBL; U46973; AAB19038.1; JOINED.  
DR EMBL; U46974; AAB19038.1; JOINED.  
DR EMBL; U46975; AAB19038.1; JOINED.  
DR EMBL; U46976; AAB19038.1; JOINED.  
DR EMBL; U46977; AAB19038.1; JOINED.  
DR EMBL; U46978; AAB19038.1; JOINED.  
DR EMBL; U46979; AAB19039.1; JOINED.  
DR EMBL; U46980; AAB19039.1; JOINED.  
DR EMBL; U46981; AAB19039.1; JOINED.  
DR EMBL; U46982; AAB19039.1; JOINED.  
DR EMBL; U46983; AAB19039.1; JOINED.  
DR EMBL; U46984; AAB19039.1; JOINED.  
DR EMBL; U46985; AAB19039.1; JOINED.  
DR EMBL; U46986; AAB19039.1; JOINED.  
DR EMBL; U46987; AAB19039.1; JOINED.  
DR EMBL; U46988; AAB19039.1; JOINED.  
DR EMBL; U46989; AAB19039.1; JOINED.  
DR EMBL; U46990; AAB19039.1; JOINED.  
DR EMBL; U46991; AAB19039.1; JOINED.  
DR EMBL; U46992; AAB19039.1; JOINED.  
DR EMBL; U46993; AAB19038.1; JOINED.  
DR EMBL; U46994; AAB19038.1; JOINED.  
DR EMBL; U46995; AAB19038.1; JOINED.  
DR EMBL; U46996; AAB19038.1; JOINED.  
DR EMBL; U46997; AAB19038.1; JOINED.  
DR EMBL; U46998; AAB19038.1; JOINED.  
DR EMBL; U46999; AAB19038.1; JOINED.  
DR EMBL; U47000; AAB19039.1; JOINED.  
DR EMBL; U47001; AAB19039.1; JOINED.  
DR EMBL; U47002; AAB19039.1; JOINED.  
DR EMBL; U47003; AAB19039.1; JOINED.  
DR EMBL; U47004; AAB19039.1; JOINED.  
DR EMBL; U47005; AAB19039.1; JOINED.  
DR EMBL; U47006; AAB19039.1; JOINED.  
DR EMBL; U47007; AAB19039.1; JOINED.  
DR EMBL; U47008; AAB19039.1; JOINED.  
DR EMBL; U47009; AAB19039.1; JOINED.  
DR EMBL; U47010; AAB19039.1; JOINED.  
DR EMBL; U47011; AAB19039.1; JOINED.  
DR EMBL; U47012; AAB19039.1; JOINED.  
DR EMBL; U47013; AAB19039.1; JOINED.  
DR EMBL; U47014; AAB19039.1; JOINED.  
DR EMBL; U47015; AAB19039.1; JOINED.  
DR EMBL; U47016; AAB19039.1; JOINED.  
DR EMBL; U47017; AAB19039.1; JOINED.  
DR EMBL; U47018; AAB19039.1; JOINED.  
DR EMBL; U47019; AAB19039.1; JOINED.  
DR EMBL; U47020; AAB19039.1; JOINED.  
DR EMBL; U47021; AAB19039.1; JOINED.  
DR EMBL; U47022; AAB19039.1; JOINED.  
DR EMBL; U47023; AAB19039.1; JOINED.  
DR EMBL; U47024; AAB19039.1; JOINED.  
DR EMBL; U47025; AAB19039.1; JOINED.  
DR EMBL; U47026; AAB19039.1; JOINED.  
DR EMBL; U47027; AAB19039.1; JOINED.  
DR EMBL; U47028; AAB19039.1; JOINED.  
DR EMBL; U47029; AAB19039.1; JOINED.  
DR EMBL; U47030; AAB19039.1; JOINED.  
DR EMBL; U47031; AAB19039.1; JOINED.  
DR EMBL; U47032; AAB19039.1; JOINED.  
DR EMBL; U47033; AAB19039.1; JOINED.  
DR EMBL; U47034; AAB19039.1; JOINED.  
DR EMBL; U47035; AAB19039.1; JOINED.  
DR EMBL; U47036; AAB19039.1; JOINED.  
DR EMBL; U47037; AAB19039.1; JOINED.  
DR EMBL; U47038; AAB19039.1; JOINED.  
DR EMBL; U47039; AAB19039.1; JOINED.  
DR EMBL; U47040; AAB19039.1; JOINED.  
DR EMBL; U47041; AAB19039.1; JOINED.  
DR EMBL; U47042; AAB19039.1; JOINED.  
DR EMBL; U47043; AAB19039.1; JOINED.  
DR EMBL; U47044; AAB19039.1; JOINED.  
DR EMBL; U47045; AAB19039.1; JOINED.  
DR EMBL; U47046; AAB19039.1; JOINED.  
DR EMBL; U47047; AAB19039.1; JOINED.  
DR EMBL; U47048; AAB19039.1; JOINED.  
DR EMBL; U47049; AAB19039.1; JOINED.  
DR EMBL; U47050; AAB19039.1; JOINED.  
DR EMBL; U47051; AAB19039.1; JOINED.  
DR EMBL; U47052; AAB19039.1; JOINED.  
DR EMBL; U47053; AAB19039.1; JOINED.  
DR EMBL; U47054; AAB19039.1; JOINED.  
DR EMBL; U47055; AAB19039.1; JOINED.  
DR EMBL; U47056; AAB19039.1; JOINED.  
DR EMBL; U47057; AAB19039.1; JOINED.  
DR EMBL; U47058; AAB19039.1; JOINED.  
DR EMBL; U47059; AAB19039.1; JOINED.  
DR EMBL; U47060; AAB19039.1; JOINED.  
DR EMBL; U47061; AAB19039.1; JOINED.  
DR EMBL; U47062; AAB19039.1; JOINED.  
DR EMBL; U47063; AAB19039.1; JOINED.  
DR EMBL; U47064; AAB19039.1; JOINED.  
DR EMBL; U47065; AAB19039.1; JOINED.  
DR EMBL; U47066; AAB19039.1; JOINED.  
DR EMBL; U47067; AAB19039.1; JOINED.  
DR EMBL; U47068; AAB19039.1; JOINED.  
DR EMBL; U47069; AAB19039.1; JOINED.  
DR EMBL; U47070; AAB19039.1; JOINED.  
DR EMBL; U47071; AAB19039.1; JOINED.  
DR EMBL; U47072; AAB19039.1; JOINED.  
DR EMBL; U47073; AAB19039.1; JOINED.  
DR EMBL; U47074; AAB19039.1; JOINED.  
DR EMBL; U47075; AAB19039.1; JOINED.  
DR EMBL; U47076; AAB19039.1; JOINED.  
DR EMBL; U47077; AAB19039.1; JOINED.  
DR EMBL; U47078; AAB19039.1; JOINED.  
DR EMBL; U47079; AAB19039.1; JOINED.  
DR EMBL; U47080; AAB19039.1; JOINED.  
DR EMBL; U47081; AAB19039.1; JOINED.  
DR EMBL; U47082; AAB19039.1; JOINED.  
DR EMBL; U47083; AAB19039.1; JOINED.  
DR EMBL; U47084; AAB19039.1; JOINED.  
DR EMBL; U47085; AAB19039.1; JOINED.  
DR EMBL; U47086; AAB19039.1; JOINED.  
DR EMBL; U47087; AAB19039.1; JOINED.  
DR EMBL; U47088; AAB19039.1; JOINED.  
DR EMBL; U47089; AAB19039.1; JOINED.  
DR EMBL; U47090; AAB19039.1; JOINED.  
DR EMBL; U47091; AAB19039.1; JOINED.  
DR EMBL; U47092; AAB19039.1; JOINED.  
DR EMBL; U47093; AAB19039.1; JOINED.  
DR EMBL; U47094; AAB19039.1; JOINED.  
DR EMBL; U47095; AAB19039.1; JOINED.  
DR EMBL; U47096; AAB19039.1; JOINED.  
DR EMBL; U47097; AAB19039.1; JOINED.  
DR EMBL; U47098; AAB19039.1; JOINED.  
DR EMBL; U47099; AAB19039.1; JOINED.  
DR EMBL; U47100; AAB19039.1; JOINED.  
DR EMBL; U47101; AAB19039.1; JOINED.  
DR EMBL; U47102; AAB19039.1; JOINED.  
DR EMBL; U47103; AAB19039.1; JOINED.  
DR EMBL; U47104; AAB19039.1; JOINED.  
DR EMBL; U47105; AAB19039.1; JOINED.  
DR EMBL; U47106; AAB19039.1; JOINED.  
DR EMBL; U47107; AAB19039.1; JOINED.  
DR EMBL; U47108; AAB19039.1; JOINED.  
DR EMBL; U47109; AAB19039.1; JOINED.  
DR EMBL; U47110; AAB19039.1; JOINED.  
DR EMBL; U47111; AAB19039.1; JOINED.  
DR EMBL; U47112; AAB19039.1; JOINED.  
DR EMBL; U47113; AAB19039.1; JOINED.  
DR EMBL; U47114; AAB19039.1; JOINED.  
DR EMBL; U47115; AAB19039.1; JOINED.  
DR EMBL; U47116; AAB19039.1; JOINED.  
DR EMBL; U47117; AAB19039.1; JOINED.  
DR EMBL; U47118; AAB19039.1; JOINED.  
DR EMBL; U47119; AAB19039.1; JOINED.  
DR EMBL; U47120; AAB19039.1; JOINED.  
DR EMBL; U47121; AAB19039.1; JOINED.  
DR EMBL; U47122; AAB19039.1; JOINED.  
DR EMBL; U47123; AAB19039.1; JOINED.  
DR EMBL; U47124; AAB19039.1; JOINED.  
DR EMBL; U47125; AAB19039.1; JOINED.  
DR EMBL; U47126; AAB19039.1; JOINED.  
DR EMBL; U47127; AAB19039.1; JOINED.  
DR EMBL; U47128; AAB19039.1; JOINED.  
DR EMBL; U47129; AAB19039.1; JOINED.  
DR EMBL; U47130; AAB19039.1; JOINED.  
DR EMBL; U47131; AAB19039.1; JOINED.  
DR EMBL; U47132; AAB19039.1; JOINED.  
DR EMBL; U47133; AAB19039.1; JOINED.  
DR EMBL; U47134; AAB19039.1; JOINED.  
DR EMBL; U47135; AAB19039.1; JOINED.  
DR EMBL; U47136; AAB19039.1; JOINED.  
DR EMBL; U47137; AAB19039.1; JOINED.  
DR EMBL; U47138; AAB19039.1; JOINED.  
DR EMBL; U47139; AAB19039.1; JOINED.  
DR EMBL; U47140; AAB19039.1; JOINED.  
DR EMBL; U47141; AAB19039.1; JOINED.  
DR EMBL; U47142; AAB19039.1; JOINED.  
DR EMBL; U47143; AAB19039.1; JOINED.  
DR EMBL; U47144; AAB19039.1; JOINED.  
DR EMBL; U47145; AAB19039.1; JOINED.  
DR EMBL; U47146; AAB19039.1; JOINED.  
DR EMBL; U47147; AAB19039.1; JOINED.  
DR EMBL; U47148; AAB19039.1; JOINED.  
DR EMBL; U47149; AAB19039.1; JOINED.  
DR EMBL; U47150; AAB19039.1; JOINED.  
DR EMBL; U47151; AAB19039.1; JOINED.  
DR EMBL; U47152; AAB19039.1; JOINED.  
DR EMBL; U47153; AAB19039.1; JOINED.  
DR EMBL; U47154; AAB19039.1; JOINED.  
DR EMBL; U47155; AAB19039.1; JOINED.  
DR EMBL; U47156; AAB19039.1; JOINED.  
DR EMBL; U47157; AAB19039.1; JOINED.  
DR EMBL; U47158; AAB19039.1; JOINED.  
DR EMBL; U47159; AAB19039.1; JOINED.  
DR EMBL; U47160; AAB19039.1; JOINED.  
DR EMBL; U47161; AAB19039.1; JOINED.  
DR EMBL; U47162; AAB19039.1; JOINED.  
DR EMBL; U47163; AAB19039.1; JOINED.  
DR EMBL; U47164; AAB19039.1; JOINED.  
DR EMBL; U47165; AAB19039.1; JOINED.  
DR EMBL; U47166; AAB19039.1; JOINED.  
DR EMBL; U47167; AAB19039.1; JOINED.  
DR EMBL; U47168; AAB19039.1; JOINED.  
DR EMBL; U47169; AAB19039.1; JOINED.  
DR EMBL; U47170; AAB19039.1; JOINED.  
DR EMBL; U47171; AAB19039.1; JOINED.  
DR EMBL; U47172; AAB19039.1; JOINED.  
DR EMBL; U47173; AAB19039.1; JOINED.  
DR EMBL; U47174; AAB19039.1; JOINED.  
DR EMBL; U47175; AAB19039.1; JOINED.  
DR EMBL; U47176; AAB19039.1; JOINED.  
DR EMBL; U47177; AAB19039.1; JOINED.  
DR EMBL; U47178; AAB19039.1; JOINED.  
DR EMBL; U47179; AAB19039.1; JOINED.  
DR EMBL; U47180; AAB19039.1; JOINED.  
DR EMBL; U47181; AAB19039.1; JOINED.  
DR EMBL; U47182; AAB19039.1; JOINED.  
DR EMBL; U47183; AAB19039.1; JOINED.  
DR EMBL; U47184; AAB19039.1; JOINED.  
DR EMBL; U47185; AAB19039.1; JOINED.  
DR EMBL; U47186; AAB19039.1; JOINED.  
DR EMBL; U47187; AAB19039.1; JOINED.  
DR EMBL; U47188; AAB19039.1; JOINED.  
DR EMBL; U47189; AAB19039.1; JOINED.  
DR EMBL; U47190; AAB19039.1; JOINED.  
DR EMBL; U47191; AAB19039.1; JOINED.  
DR EMBL; U47192; AAB19039.1; JOINED.  
DR EMBL; U47193; AAB19039.1; JOINED.  
DR EMBL; U47194; AAB19039.1; JOINED.  
DR EMBL; U47195; AAB19039.1; JOINED.  
DR EMBL; U47196; AAB19039.1; JOINED.  
DR EMBL; U47197; AAB19039.1; JOINED.  
DR EMBL; U47198; AAB19039.1; JOINED.  
DR EMBL; U47199; AAB19039.1; JOINED.  
DR EMBL; U47200; AAB19039.1; JOINED.  
DR EMBL; U47201; AAB19039.1; JOINED.  
DR EMBL; U47202; AAB19039.1; JOINED.  
DR EMBL; U47203; AAB19039.1; JOINED.  
DR EMBL; U47204; AAB19039.1; JOINED.  
DR EMBL; U47205; AAB19039.1; JOINED.  
DR EMBL; U47206; AAB19039.1; JOINED.  
DR EMBL; U47207; AAB19039.1; JOINED.  
DR EMBL; U47208; AAB19039.1; JOINED.  
DR EMBL; U47209; AAB19039.1; JOINED.  
DR EMBL; U47210; AAB19039.1; JOINED.  
DR EMBL; U47211; AAB19039.1; JOINED.  
DR EMBL; U47212; AAB19039.1; JOINED.  
DR EMBL; U47213; AAB19039.1; JOINED.  
DR EMBL; U47214; AAB19039.1; JOINED.  
DR EMBL; U47215; AAB19039.1; JOINED.  
DR EMBL; U47216; AAB19039.1; JOINED.  
DR EMBL; U47217; AAB19039.1; JOINED.  
DR EMBL; U47218; AAB19039.1; JOINED.  
DR EMBL; U47219; AAB19039.1; JOINED.  
DR EMBL; U47220; AAB19039.1; JOINED.  
DR EMBL; U47221; AAB19039.1; JOINED.  
DR EMBL; U47222; AAB19039.1; JOINED.  
DR EMBL; U47223; AAB19039.1; JOINED.  
DR EMBL; U47224; AAB19039.1; JOINED.  
DR EMBL; U47225; AAB19039.1; JOINED.  
DR EMBL; U47226; AAB19039.1; JOINED.  
DR EMBL; U47227; AAB19039.1; JOINED.  
DR EMBL; U47228; AAB19039.1; JOINED.  
DR EMBL; U47229; AAB19039.1; JOINED.  
DR EMBL; U47230; AAB19039.1; JOINED.  
DR EMBL; U47231; AAB19039.1; JOINED.  
DR EMBL; U47232; AAB19039.1; JOINED.  
DR EMBL; U47233; AAB19039.1; JOINED.  
DR EMBL; U47234; AAB19039.1; JOINED.  
DR EMBL; U47235; AAB19039.1; JOINED.  
DR EMBL; U47236; AAB19039.1; JOINED.  
DR EMBL; U47237; AAB19039.1; JOINED.  
DR EMBL; U47238; AAB19039.1; JOINED.  
DR EMBL; U47239; AAB19039.1; JOINED.  
DR EMBL; U47240; AAB19039.1; JOINED.  
DR EMBL; U47241; AAB19039.1; JOINED.  
DR EMBL; U47242; AAB19039.1; JOINED.  
DR EMBL; U47243; AAB19039.1; JOINED.  
DR EMBL; U47244; AAB19039.1; JOINED.  
DR EMBL; U47245; AAB19039.1; JOINED.  
DR EMBL; U47246; AAB19039.1; JOINED.  
DR EMBL; U47247; AAB19039.1; JOINED.  
DR EMBL; U47248; AAB19039.1; JOINED.  
DR EMBL; U47249; AAB19039.1; JOINED.  
DR EMBL; U47250; AAB19039.1; JOINED.  
DR EMBL; U47251; AAB19039.1; JOINED.  
DR EMBL; U47252; AAB19039.1; JOINED.  
DR EMBL; U47253; AAB19039.1; JOINED.  
DR EMBL; U47254; AAB19039.1; JOINED.  
DR EMBL; U47255; AAB19039.1; JOINED.  
DR EMBL; U47256; AAB19039.1; JOINED.  
DR EMBL; U47257; AAB19039.1; JOINED.  
DR EMBL; U47258; AAB19039.1; JOINED.  
DR EMBL; U47259; AAB19039.1; JOINED.  
DR EMBL; U47260; AAB19039.1; JOINED.  
DR EMBL; U47261; AAB19039.1; JOINED.  
DR EMBL; U47262; AAB19039.1; JOINED.  
DR EMBL; U47263; AAB19039.1; JOINED.  
DR EMBL; U47264; AAB19039.1; JOINED.  
DR EMBL; U47265; AAB19039.1; JOINED.  
DR EMBL; U47266; AAB19039.1; JOINED.  
DR EMBL; U47267; AAB19039.1; JOINED.  
DR EMBL; U47268; AAB19039.1; JOINED.  
DR EMBL; U47269; AAB19039.1; JOINED.  
DR EMBL; U47270; AAB19039.1; JOINED.  
DR EMBL; U47271; AAB19039.1; JOINED.  
DR EMBL; U47272; AAB19039.1; JOINED.  
DR EMBL; U47273; AAB19039.1; JOINED.  
DR EMBL; U47274; AAB19039.1; JOINED.  
DR EMBL; U47275; AAB19039.1; JOINED.  
DR EMBL; U47276; AAB19039.1; JOINED.  
DR EMBL; U47277; AAB19039.1; JOINED.  
DR EMBL; U47278; AAB19039.1; JOINED.  
DR EMBL; U47279; AAB19039.1; JOINED.  
DR EMBL; U47280; AAB19039.1; JOINED.  
DR EMBL; U47281; AAB19039.1; JOINED.  
DR EMBL; U47282; AAB19039.1; JOINED.  
DR EMBL; U47283; AAB19039.1; JOINED.  
DR EMBL; U47284; AAB19039.1; JOINED.  
DR EMBL; U47285; AAB19039.1; JOINED.  
DR EMBL; U47286; AAB19039.1; JOINED.  
DR EMBL; U47287; AAB19039.1; JOINED.  
DR EMBL; U47288; AAB19039.1; JOINED.  
DR EMBL; U47289; AAB19039.1; JOINED.  
DR EMBL; U47290; AAB19039.1; JOINED.  
DR EMBL; U47291; AAB19039.1; JOINED.  
DR EMBL; U47292; AAB19039.1; JOINED.  
DR EMBL; U47293; AAB19039.1; JOINED.  
DR EMBL; U47294; AAB19039.1; JOINED.  
DR EMBL; U47295; AAB19039.1; JOINED.  
DR EMBL; U47296; AAB19039.1; JOINED.  
DR EMBL; U47297; AAB19039.1; JOINED.  
DR EMBL; U47298; AAB19039.1; JOINED.  
DR EMBL; U47299; AAB19039.1; JOINED.  
DR EMBL; U47300; AAB19039.1; JOINED.  
DR EMBL; U47301; AAB19039.1; JOINED.  
DR EMBL; U47302; AAB19039.1; JOINED.  
DR EMBL; U47303; AAB19039.1; JOINED.  
DR EMBL; U47304; AAB19039.1; JOINED.  
DR EMBL; U47305; AAB19039.1; JOINED.  
DR EMBL; U47306; AAB19039.1; JOINED.  
DR EMBL; U47307; AAB19039.1; JOINED.  
DR EMBL; U47308; AAB19039.1; JOINED.  
DR EMBL; U47309; AAB19039.1; JOINED.  
DR EMBL; U47310; AAB19039.1; JOINED.  
DR EMBL; U47311; AAB19039.1; JOINED.  
DR EMBL; U47312; AAB19039.1; JOINED.  
DR EMBL; U47313; AAB19039.1; JOINED.  
DR EMBL; U47314; AAB19039.1; JOINED.  
DR EMBL; U47315; AAB19039.1; JOINED.  
DR EMBL; U47316; AAB19039.1; JOINED.  
DR EMBL; U47317; AAB19039.1; JOINED.  
DR EMBL; U47318; AAB19039.1; JOINED.  
DR EMBL; U47319; AAB19039.1; JOINED.  
DR EMBL; U47320; AAB19039.1; JOINED.  
DR EMBL; U47321; AAB19039.1; JOINED.  
DR EMBL; U47322; AAB19039.1; JOINED.  
DR EMBL; U47323; AAB19039.1; JOINED.  
DR EMBL; U47324; AAB19039.1; JOINED.  
DR EMBL; U47325; AAB19039.1; JOINED.  
DR EMBL; U47326; AAB19039.1; JOINED.  
DR EMBL; U47327; AAB19039.1; JOINED.  
DR EMBL; U47328; AAB19039.1; JOINED.  
DR EMBL; U47329; AAB19039.1; JOINED.  
DR EMBL; U47330; AAB19039.1; JOINED.  
DR EMBL; U47331; AAB19039.1; JOINED.  
DR EMBL; U47332; AAB19039.1; JOINED.  
DR EMBL; U47333; AAB19039.1; JOINED.  
DR EMBL; U47334; AAB19039.1; JOINED.  
DR EMBL; U47335; AAB19039.1; JOINED.  
DR EMBL; U47336; AAB19039.1; JOINED.  
DR EMBL; U47337; AAB19039.1; JOINED.  
DR EMBL; U47338; AAB19039.1; JOINED.  
DR EMBL; U47339; AAB19039.1; JOINED.  
DR EMBL; U47340; AAB19039.1; JOINED.  
DR EMBL; U47341; AAB19039.1; JOINED.  
DR EMBL; U47342; AAB19039.1; JOINED.  
DR EMBL; U47343; AAB19039.1; JOINED.  
DR EMBL; U47344; AAB19039.1; JOINED.  
DR EMBL; U47345; AAB19039.1; JOINED.  
DR EMBL; U47346; AAB19039.1; JOINED.  
DR EMBL; U47347; AAB19039.1; JOINED.  
DR EMBL; U47348; AAB19039.1; JOINED.  
DR EMBL; U47349; AAB19039.1; JOINED.  
DR EMBL; U47350; AAB19039.1; JOINED.  
DR EMBL; U47351; AAB19039.1; JOINED.  
DR EMBL; U47352; AAB19039.1; JOINED.  
DR EMBL; U47353; AAB19039.1; JOINED.  
DR EMBL; U47354; AAB19039.1; JOINED.  
DR EMBL; U47355; AAB19039.1; JOINED.  
DR EMBL; U47356; AAB19039.1; JOINED.  
DR EMBL; U47357; AAB19039.1; JOINED.  
DR EMBL; U47358; AAB19039.1; JOINED.  
DR EMBL; U47359; AAB19039.1; JOINED.  
DR EMBL; U47360; AAB19039.1; JOINED.  
DR EMBL; U47361; AAB19039.1; JOINED.  
DR EMBL; U47362; AAB19039.1; JOINED.  
DR EMBL; U47363; AAB19039.1; JOINED.  
DR EMBL; U47364; AAB19039.1; JOINED.  
DR EMBL; U47365; AAB19039.1; JOINED.  
DR EMBL; U47366; AAB19039.1; JOINED.  
DR EMBL; U47367; AAB19039.1; JOINED.  
DR EMBL; U47368; AAB19039.1; JOINED.  
DR EMBL; U47369; AAB19039.1; JOINED.  
DR EMBL; U47370; AAB19039.1; JOINED.  
DR EMBL; U47371; AAB19039.1; JOINED.  
DR EMBL; U47372; AAB19039.1; JOINED.  
DR EMBL; U47373; AAB19039.1; JOINED.  
DR EMBL; U47374; AAB19039.1; JOINED.  
DR EMBL; U47375; AAB19039.1; JOINED.  
DR EMBL; U47376; AAB19039.1; JOINED.  
DR EMBL; U47377; AAB19

DR EMBL; L22763; AA16338.1; -.  
DR PIR; A54122; CGHUGB.  
DR Genew; HGNC:2208; COL4A6.  
DR MIM; 303631; -.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.  
DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 23.  
DR ProDom; PD000007; Clg\_helix; 4.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR ProDom; PD00111; C4; 2.  
DR SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal;  
KW Alternative splicing; Polymorphism.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 1691 COLLAGEN ALPHA 6 (IV) CHAIN.  
FT DOMAIN 23 46 7S DOMAIN.  
Query Match 3.3%; Score 8; DB 1; Length 1691;  
Best/Local Similarity 100.0%; Pred. No. 9.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 234 SRCQVCMK 241  
Db 1682 SRCQVCMK 1689  
|||||||  
RESULT 13  
CA14 DROME STANDARD; PRT; 1775 AA.  
AC P08120;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Collagen alpha 1(IV) chain precursor.  
GN CG25C OR DCG1.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89054012; PubMed=3142875;  
RA Blumberg B., Mackrell A.J., Fessler J.H.;  
RT "Drosophila basement membrane procollagen alpha 1(IV). II. Complete  
RT cDNA sequence, genomic structure, and general implications for  
RT supramolecular assemblies."  
RL J. Biol. Chem. 263:18328-18337(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Blumberg B.;  
RA Thesis (1987), University of California / Los Angeles, U.S.A.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Mackrell A.J.;  
RN Thesis (1992), University of California / Los Angeles, U.S.A.  
RP SEQUENCE OF 1065-1775 FROM N.A.  
RX MEDLINE=87194801; PubMed=3106346;  
RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,  
RA Natzie J.E., Fessler J.H.;  
RT "Basement membrane procollagen IV and its specialized carboxyl domain  
RT are conserved in Drosophila, mouse, and human."  
RL J. Biol. Chem. 262:5947-5950(1987).  
RN [5]  
RP SEQUENCE OF 1355-1775 FROM N.A.  
RX MEDLINE=87246644; PubMed=3109906;  
RA Cecchini J.P., Knibiehler B., Mirre C., le Parco Y.;  
RT "Evidence for a type-IV-related collagen in Drosophila melanogaster.  
RT Evolutionary constancy of the carboxyl-terminal noncollagenous  
RT domain.";  
RL Eur. J. Biochem. 165:587-593(1987).  
RN [6]  
RP SEQUENCE OF 762-1230 FROM N.A.  
RX MEDLINE=82197577; PubMed=6210912;  
RA Monson J.M., Natzie J., Friedman J., McCarthy B.J.;  
RA "Expression and novel structure of a collagen gene in Drosophila.";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).  
CC -|- FUNCTION: Collagen type IV is specific for basement membranes.  
CC -|- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.  
CC Type IV collagen forms a mesh-like network linked through  
CC intermolecular interactions between 7S domains and between NC1  
CC domains.  
CC -|- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the G-  
CC X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -|- PM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -|- PM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M23704; AAA28404.1; -.  
CC EMBL; M96575; AAB55184.1; -.  
CC EMBL; J02727; AAA28423.1; -.  
CC EMBL; M28334; AAA28422.1; -.  
CC EMBL; V00200; CAA23486.2; -.  
CC PIR; A31893; A31893.  
CC FlyBase; FBgn000299; Cg25C.  
CC GO; GO:0005587; C:collagen type IV; NAS.  
CC InterPro; IPR008161; Clg\_helix.  
CC InterPro; IPR008160; Collagen.  
CC InterPro; IPR001442; Procollagn4\_C.  
CC Pfam; PF01413; C4; 2.  
CC Pfam; PF01391; Collagen; 25.  
CC ProDom; PD000007; Clg\_helix; 9.  
CC ProDom; PD003923; ProcollagnC4; 1.  
CC SMART; SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 23  
FT PROPEP 24 ? AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN ? 1775 COLLAGEN ALPHA 1(IV) CHAIN.  
FT DOMAIN ? 1544 TRIPLE-HELICAL REGION.  
FT NC1 1545 1775 NONHELICAL REGION (NC1).  
FT DISULFID 1569 1655 OR 1652 (BY SIMILARITY).  
FT DISULFID 1599 1652 OR 1655 (BY SIMILARITY).  
FT DISULFID 1611 1617 BY SIMILARITY.  
FT DISULFID 1674 1770 OR 1767 (BY SIMILARITY).  
FT DISULFID 1708 1767 OR 1770 (BY SIMILARITY).  
FT DISULFID 1720 1727 BY SIMILARITY.  
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (PROBABLE).  
FT CONFLICT 948 948 L -> S (IN REF. 6).  
FT CONFLICT 997 997 S -> T (IN REF. 6).  
FT CONFLICT 1357 1357 Q -> K (IN REF. 5).  
FT CONFLICT 1360 1360 Q -> K (IN REF. 5).  
FT CONFLICT 1373 1373 T -> I (IN REF. 5).  
FT CONFLICT 1496 1496 L -> R (IN REF. 5).  
FT CONFLICT 1507 1511 ETGNNV -> RAGOR (IN REF. 5).  
FT CONFLICT 1529 1529 E -> K (IN REF. 5).

```
FT CONFLICT 1733 1733 M -> I (IN REF. 5).
SQ SEQUENCE 1775 AA; 174119 MW; 2DE5AB23149525CD CRC64;

Query Match 3.3%; Score 8; DB 1; Length 1775;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
DB 1765 SRCQVCWK 1772
|||||

RESULT 14
VB03_VACCC STANDARD; PRT; 124 AA.
AC P21050;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein B3.
GN B3R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP004599; BAC13963.1; -.
DR HAMAP, MF_00651; -.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; YqgFc; 1.
DR SMART; SMC0732; YqgFc; 1.
DR TIGRFAMs; TIGR00250; TIGR00250; 1.
DR HydroLase; Nuclease; DNA repair; DNA recombination; Complete proteome.
KW SEQUENCE 137 AA; 15239 MW; 500BDADCFCBDB7CEC CRC64;
SQ SEQUENCE 137 AA; 15239 MW; 500BDADCFCBDB7CEC CRC64;

Query Match 2.9%; Score 7; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIS 234
DB 44 ELEKIIS 50
|||||

RESULT 16
RNB_HSV2H STANDARD; PRT; 151 AA.
AC P89479;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Potential RNA-binding protein.
GN US11.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z86099; CAB06719.1; -.
DR DNA-binding; RNA-binding; Repeat; Nuclear protein.
FT DOMAIN 90 146 11 X 6 AA TANDEM REPEATS.
FT REPEAT 90 95 1.
FT REPEAT 96 101 2.

QY 219 PIPSTVK 225
DB 38 PIPSTVK 44
|||||

RESULT 15
RUVX_OCEIH STANDARD; PRT; 137 AA.
AC O8EPT1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.-.-).
GN OB2007.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=162710;
RN [1]
```



FT REPEAT 102 104 3.  
FT REPEAT 105 110 4.  
FT REPEAT 111 116 5.  
FT REPEAT 117 122 6.  
FT REPEAT 123 128 7.  
FT REPEAT 129 130 8.  
FT REPEAT 131 134 9.  
FT REPEAT 135 140 10.  
FT REPEAT 141 146 11.  
SQ SEQUENCE 151 AA; 16297 MW; FAB751F23C3DB6AE CRC64;

Query Match 2.9%; Score 7; DB 1; Length 151;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GKRGSQ 10

DB 61 GKRGSQ 67

## RESULT 17

NUPM\_NEUCR

ID NUPM\_NEUCR STANDARD; PRT; 183 AA.

AC P21976;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE NADH-ubiquinone oxidoreductase 20.8 kDa subunit (EC 1.6.5.3)

DE (EC 1.6.5.3.3)

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-45.

RA MEDLINE=90330847; PubMed=2142943;

RA Videira A., Tropisch M., Wachter E., Schneider H., Werner S.;

RT "Molecular cloning of subunits of complex I from Neurospora crassa.

RT Primary structure and in vitro expression of a 22-kDa polypeptide.";

RL J. Biol. Chem. 265:13060-13065 (1990).

CC -!- FUNCTION: Transfer of electrons from NADH to the respiratory

CC chain. The immediate electron acceptor for the enzyme is believed

CC to be ubiquinone.

CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.

CC -!- COFACTOR: Binds 1 iron-sulfur cluster (potential).

CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.

CC THIS IS A COMPONENT OF THE HYDROPHOBIC FRACTION.

CC -!- SUBCELLULAR LOCATION: Matrix and cytoplasmic side of the

CC mitochondrial inner membrane.

CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 19 kDa SUBUNIT FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M55323; AAA33571.1; -

DR PIR; T47251; T47251.

DR InterPro; IPR008697; NDUFA8.

DR Pfam; PF05850; NDUFA8; 1.

KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.

SQ SEQUENCE 183 AA; 20911 MW; A2574693F41093D4 CRC64;

## Query Match

Best Local Similarity 2.9%; Score 7; DB 1; Length 183;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 SCLEEF 183

DB 177 SCLEEF 183

Db 90 SCLEEF 96

## RESULT 18

SODM\_PROF

ID SODM\_PROF STANDARD; PRT; 201 AA.

AC P80239;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).

GN SODA.

OS Propionibacterium freudenreichii shermanii.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.

OC NCBI\_TaxID=1752;

RN [1]

RP SEQUENCE.

RC STRAIN=P23;

RE MEDLINE=94139724; PubMed=8307013;

RA Meier B., Sehn A.P., Schinina M.E., Barra D.;

RT "In vivo incorporation of copper into the iron-exchangeable and

RT manganese-exchangeable superoxide dismutase from Propionibacterium

RT shermanii. Amino acid sequence and identity of the protein

RT moieties.";

RL Eur. J. Biochem. 219:463-468 (1994).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RA Schmidt M., Meier B., Parak F.;

RT "X-ray structure of the cambialistic superoxide dismutase from

RT Propionibacterium shermanii active with Fe or Mn.";

RL J. Biol. Inorg. Chem. 1:532-541 (1996).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).

RA Schmidt M., Scherx C., Iakovleva O., Nolling H.F., Meier B., Parak F.;

RT Submitted (SEP-1997) to the PDB data bank.

RN [4]

RP X-RAY CRYSTALLOGRAPHY (1.35 ANGSTROMS).

RE MEDLINE=99248073; PubMed=10231372;

RA Schmidt M.;

RT "Manipulating the coordination number of the ferric iron within the

RT cambialistic superoxide dismutase of Propionibacterium shermanii by

RT changing the pH-value. A crystallographic analysis.";

RL Eur. J. Biochem. 262:117-127 (1999).

CC -!- FUNCTION: Destroys radicals which are normally produced within the

CC cells and which are toxic to biological systems.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese or iron ion per subunit.

CC -!- SUBUNIT: Homotetramer.

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase

CC family.

DR PIR; JC4396; JC4396.

DR PDB; 1AR4; 12-NOV-97.

DR PDB; 1AR5; 12-NOV-97.

DR PDB; 1AVN; 18-MAR-98.

DR PDB; 1BS3; 15-JUN-99.

DR PDB; 1BSM; 15-JUN-99.

DR PDB; 1B78; 15-JUN-99.

DR InterPro; IPR001189; SODismutase.

DR Pfam; PF00081; sodef; 1.

DR Pfam; PF02777; sodef; 1.

DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom; PD000475; SODismutase; 1.

DR PROSITE; PS00088; SOD MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Iron; 3D-structure.

FT METAL 27 27

FT METAL 75 75

FT METAL 161 161

FT METAL 165 165

FT TURN 12 18

FT HELIX 21 29

FT TURN 30 30

FT TURN 31 52

FT HELIX 31 52

```
FT TURN 53 53
FT TURN 56 57
FT TURN 58 80
FT TURN 81 81
FT STRAND 82 82
FT TURN 85 86
FT TURN 86 86
FT HELIX 94 104
FT HELIX 107 119
FT TURN 120 120
FT STRAND 125 132
FT TURN 133 136
FT STRAND 137 144
FT TURN 145 147
FT STRAND 148 148
FT TURN 152 153
FT STRAND 155 161
FT HELIX 164 166
FT TURN 167 167
FT HELIX 168 171
FT TURN 172 173
FT HELIX 175 182
FT STRAND 183 185
FT STRAND 186 186
FT HELIX 188 198
FT TURN 199 199
SQ SEQUENCE 201 AA; 22633 MW; 5BFEF424C7B32E00 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 14 ALEPYIS 20

RESULT 19
SODF_METTM STANDARD; PRT; 202 AA.
AC Q60036;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD.
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=79929;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95301176; PubMed=7781971;
RA Meile L., Fischer K., Leisinger T.;
RT "Characterization of the superoxide dismutase gene and its upstream
region from Methanobacterium thermoautotrophicum Marburg.";
RL FEMS Microbiol. Lett. 128:247-253 (1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit (By similarity).
CC -!- SUBUNIT: Homotetramer.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74264; CAA52323.1; -.
```

```
DR PIR; S51097; S51097.
DR HSSP; Q08713; 1B06.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase; Metal-binding; Iron.
FT METAL 30 30 IRON (BY SIMILARITY).
FT METAL 78 78 IRON (BY SIMILARITY).
FT METAL 164 164 IRON (BY SIMILARITY).
FT METAL 168 168 IRON (BY SIMILARITY).
SQ SEQUENCE 202 AA; 23829 MW; SC4FBE27EE63223 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 17 ALEPYIS 23

RESULT 20
SODF_SULAC STANDARD; PRT; 210 AA.
AC Q08713;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93326644; PubMed=83341170;
RA Klenk H.-P., Schleper C., Schwass V., Brudler R.;
RT "Nucleotide sequence, transcription and phylogeny of the gene
encoding the superoxide dismutase of Sulfolobus acidocaldarius.";
RL Biochim. Biophys. Acta 1174:95-98 (1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=99096955; PubMed=9878438;
RA Knapp S., Kardinahl S., Hellgren N., Tibbelin G., Schaefer G.,
RA Ladenstein R.;
RT "Refined crystal structure of a superoxide dismutase from the
hyperthermophilic archaeon Sulfolobus acidocaldarius at 2.2-A
resolution.";
RL J. Mol. Biol. 285:689-702 (1999).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit.
CC -!- SUBUNIT: HOMOTETRAMER AT HIGH TEMPERATURE; HOMODIMER AT ROOM
TEMPERATURE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63386; CAA44993.1; -.
```

DR PIR; S34616; S34616.  
 DR PDB; 1B06; 18-NOV-99.  
 DR InterPro: IPR001189; SODismutase.  
 DR Pfam; PF00081; sodfe; 1.  
 DR Pfam; PF02777; sodfe; 1.  
 DR PRINTS; PR01703; MNSODISMUTASE.  
 DR PRODOM; PD000475; SODismutase; 1.  
 DR PROSITE; PS00088; SOD\_MN; FALSE NEG.  
 KW Oxidoreductase; Metal-binding; Iron; 3D-structure.  
 FT INIT\_MET 0  
 FT METAL 33 33 IRON.  
 FT METAL 84 84 IRON.  
 FT METAL 170 170 IRON.  
 FT METAL 174 174 IRON.  
 FT TURN 18 24  
 FT HELIX 27 35  
 FT TURN 36 36  
 FT HELIX 37 57  
 FT TURN 58 63  
 FT TURN 62 63  
 FT HELIX 67 89  
 FT TURN 90 90  
 FT STRAND 91 91  
 FT TURN 94 96  
 FT HELIX 103 113  
 FT HELIX 116 129  
 FT STRAND 134 140  
 FT TURN 142 144  
 FT STRAND 147 153  
 FT TURN 154 156  
 FT STRAND 157 157  
 FT TURN 161 162  
 FT STRAND 165 170  
 FT HELIX 173 175  
 FT TURN 176 176  
 FT HELIX 177 180  
 FT TURN 181 182  
 FT HELIX 184 191  
 FT STRAND 192 194  
 FT STRAND 195 195  
 FT HELIX 197 205  
 FT TURN 206 210  
 SQ SEQUENCE 210 AA; 24135 MW; 086CCAB277D99FBB CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 210;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 114 ALPEYIS 120  
 Db 20 ALPEYIS 26  
 RESULT 21.  
 SODF\_SULSO  
 ID\_SODF\_SULSO STANDARD; PRT; 210 AA.  
 AC P08357;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Superoxide dismutase [Fe] (EC 1.15.1.1).  
 GN SOD OR SSO0316.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;  
 RX MEDLINE=98088931; PubMed=9428655;  
 RA De Russo A., Rullo R., Nitti G., Masullo M., Bocchini V.;  
 RT "Iron superoxide dismutase from the archaeon Sulfolobus solfataricus:  
 average hydrophobicity and amino acid weight are involved in the

adaptation of proteins to extreme environments."; Biochim. Biophys. Acta 1343:23-30(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99098843; PubMed=9880816;  
 RA Yamano S., Maruyama T.;  
 RT "An azide-insensitive superoxide dismutase from a hyperthermophilic  
 archaeon, Sulfolobus solfataricus." J. Biochem. 125:186-193(1999).  
 RL J. Biochem. 125:186-193(1999).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;  
 RX MEDLINE=21145482; PubMed=11248699;  
 RA De Vendittis E., Ursby T., Rullo R., Gogliettino M.A., Masullo M.,  
 Bocchini V.;  
 RT "Phenylmethanesulfonyl fluoride inactivates an archaeal superoxide  
 dismutase by chemical modification of a specific tyrosine residue:  
 cloning, sequencing and expression of the gene coding for Sulfolobus  
 solfataricus superoxide dismutase." Eur. J. Biochem. 268:1794-1801(2001).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,  
 De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,  
 Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,  
 Charlebois R.L., Doellittle W.F., Duguet M., Gaasterland T.,  
 Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=99134399; PubMed=9931259;  
 RA Urby T., Adinolfi B.S., Al-Karadaghi S., de Vendittis E.,  
 Bocchini V.;  
 RT "Iron superoxide dismutase from the archaeon Sulfolobus solfataricus:  
 analysis of structure and thermostability." J. Mol. Biol. 286:189-205(1999).  
 RL J. Mol. Biol. 286:189-205(1999).  
 CC -!- FUNCTION: Destroys radicals which are normally produced within the  
 cells and which are toxic to biological systems.  
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
 CC -!- COFACTOR: Binds 1 iron ion per subunit.  
 CC -!- SUBUNIT: Homotetramer.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase  
 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB012620; BAA75509.1; -.  
 CC EMBL; Y15326; CAA75583.1; -.  
 CC EMBL; AE006666; AAK40652.1; -.  
 CC PIR; E90174; E90174.  
 CC PDB; 1SSS; 09-APR-99.  
 CC InterPro; IPR001189; SODismutase.  
 CC Pfam; PF00081; sodfe; 1.  
 CC Pfam; PF02777; sodfe; 1.  
 CC PRINTS; PR01703; MNSODISMUTASE.  
 CC PRODOM; PD000475; SODismutase; 1.  
 CC PROSITE; PS00088; SOD\_MN; FALSE NEG.  
 KW Oxidoreductase; Metal-binding; Iron; 3D-structure; Complete proteome.  
 FT INIT\_MET 0  
 FT METAL 37 37 IRON.  
 FT METAL 84 84 IRON.

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF236110; AAF36989.1; -.
CC HSSP; P80857; ISSS.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe_C; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PROSITE; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD.MN; FALSE NEG.
CC Oxidoreductase; Metal-binding; Iron.
CC METAL 34 IRON (BY SIMILARITY).
CC FT METAL 82 IRON (BY SIMILARITY).
CC FT METAL 171 IRON (BY SIMILARITY).
CC FT METAL 175 IRON (BY SIMILARITY).
CC SQ SEQUENCE 211 AA; 24342 MW; 5A89FFB40CF77065 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 114 ALEPYIS 120
Db 21 ALEPYIS 27

RESULT 23
SODF_PYRAE STANDARD; PRT; 211 AA.
AC O93724;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Fe] (EC 1.15.1.1).
GN SOD OR PAB0274.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RA Baikov C.J., Slupska M.M., Miller J.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum".
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 iron ion per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U82371; AAD00533.2; -.
CC EMBL; AE009759; AAL62675.1; -.

```

```

DR HSP: Q08713; 1B06
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe; 1.
DR PRINTS: PF02777; sodfe; 1.
DR PRODOM: PD000475; SODismutase; 1.
DR PROSITE: PS00088; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Iron; Complete proteome.
FT METAL 31 31 IRON (BY SIMILARITY).
FT METAL 79 79 IRON (BY SIMILARITY).
FT METAL 165 165 IRON (BY SIMILARITY).
FT METAL 169 169 IRON (BY SIMILARITY).
SQ SEQUENCE 211 AA; 24204 MW; F5943D397DD31561 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 16; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 114 ALEPYIS 120
Db 18 ALEPYIS 24

RESULT 24
SODM_YEAST STANDARD; PRT; 233 AA.
AC P00447;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
GN SOD2 OR YHR008C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85127011; PubMed=3882422;
RA Marras C.A.M., van Loon A.P.G.M., Oudshoorn P., van Steeg H.,
RA Grivell L.A., Slater E.C.;
RA "Nucleotide sequence analysis of the nuclear gene coding for
RT manganese superoxide dismutase of yeast mitochondria, a gene
RT previously assumed to code for the Rieske iron-sulphur protein.";
RL Eur. J. Biochem. 147:153-161(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faville A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
RN [3]
RP SEQUENCE OF 1-39 FROM N.A.
RX MEDLINE=89211942; PubMed=3072251;
RA Schrank I.S., Sims P.F., Oliver S.G.;
RT "Functional expression of the yeast Mn-superoxide dismutase gene in
RT Escherichia coli requires deletion of the signal peptide sequence.";
RL Gene 73:121-130(1988).
RN [4]
RP SEQUENCE OF 27-233.
RA Ditlow C., Johansen J.T., Martin B.M., Svendsen I.;
RT "The complete amino acid sequence of manganese-superoxide dismutase
RT from Saccharomyces cerevisiae.";
RL Carlsberg Res. Commun. 47:81-91(1982).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
-!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
-!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
-!- SUBUNIT: Homotetramer.
-!- SUBCELLULAR LOCATION: Mitochondrial matrix.
-!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X02156; CAA26092.1; -.
EMBL; U10400; AAB68939.1; -.
EMBL; M24079; AAA35065.1; -.
PIR; A00521; DGBYN.
HSP; P04179; IABM.
GeneOnline; 139325; -.
SGD; S0001050; SOD2.
GO; GO:0005759; C:mitochondrial matrix; IDA.
GO; GO:0008383; F:manganese superoxide dismutase activity; IDA.
GO; GO:0006800; P:oxygen and reactive oxygen species metabolism; IMP.
InterPro: IPR001189; SODismutase.
Pfam; PF00081; sodfe; 1.
PRINTS; PF02777; sodfe; 1.
PRODOM; PD000475; SODismutase; 1.
PROSITE; PS00088; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Manganese; Mitochondrion;
FT TRANSIT 1 25 MITOCHONDRION
FT CHAIN 27 233 SUPEROXIDE DISMUTASE [MN].
FT METAL 52 52 MANGANESE (BY SIMILARITY).
FT METAL 107 107 MANGANESE (BY SIMILARITY).
FT METAL 194 194 MANGANESE (BY SIMILARITY).
FT METAL 198 198 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 233 AA; 25774 MW; 88A9391FBB31D06E CRC64;

Query Match 2.9%; Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 39 ALEPYIS 45

RESULT 25
SODM_NEUCR STANDARD; PRT; 245 AA.
ID SODM_NEUCR
IC Q9Y783;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
GN SOD-2 OR 18F11.030.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Dvorachek W.H., Natvig D.N.;
RT "Characterization of sod-2, the Neurospora crassa gene for manganese
RT superoxide dismutase.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=74-OR23-1A / FGSC 987;

```

```

RX MEDLINE=22542210; PubMed=12555011;
RA Nannhaupf G., Monzone C., Haase D., Mewes H.-W., Aign V.,
RA Rehsel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RL Neurospora genome sequence.";
RL Nucleic Acids Res. 31:1944-1954(2003).
CC -1- FUNCTION: Destroy radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF118809; AAD28503.1; -.
DR EMBL; AL670011; CAD21408.1; -.
DR HSSP; P04179; IABM.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodef; 1.
DR Pfam; PF02777; sodef_C; 1.
DR PRINIS; PR01703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD MN; 1.
KW Oxidoreductase; Metal-binding; Manganese; Mitochondrion;
KW Transit peptide.
FT TRANSIT 1 32 MITOCHONDRION (POTENTIAL).
FT CHAIN 33 245 SUPEROXIDE DISMUTASE [MN].
FT METAL 58 58 MANGANESE (BY SIMILARITY).
FT METAL 106 106 MANGANESE (BY SIMILARITY).
FT METAL 196 196 MANGANESE (BY SIMILARITY).
FT METAL 200 200 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 245 AA; 27019 MW; FF288947F57676AD CRC64;

Query Match 2.9%; Score 7; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120
Db 45 ALPEYIS 51
|||||

RESULT 26
YACF SALTY
ID YACF SALTY STANDARD; PRT; 247 AA.
AC Q8X8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0289 protein yacf.
GN YACF OR STM0139 OR STY0161 OR T0145.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21514948; PubMed=1177609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
(2)
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=1177608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Mcule S., O'Garra P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
(3)
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- SIMILARITY: Belongs to the UPF0289 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AE008700; AAL19103.1; -.
DR EMBL; AL627265; CAD01298.1; -.
DR EMBL; AE016834; AAO67877.1; -.
DR StyGene; SG????; yacf.
DR HAMAP; MF_01092; -.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 247 AA; 28425 MW; E1B9826AD004BE48 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212
Db 157 WLASLNP 163
|||||

RESULT 27
CFAD MOUSE
ID CFAD MOUSE STANDARD; PRT; 259 AA.
AC P03953; Q61280;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Complement factor D precursor (EC 3.4.21.46) (C3 convertase activator)
DE (Properdin factor D) (Adipsin) (28 kDa protein, adipocyte).
GN DF OR ADN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=86278164; PubMed=3015943;
```





```

DR InterPro: IPR003000; SIR2.
DR Pfam: PF02146; SIR2; 1.
DR PROSITE: PS50305; SIRTUIN; 1.
KW Hydrolase; NAD; transcription regulation; Repressor; Nuclear protein;
FT METAL-binding; Zinc.
KW DOMAIN 83 370 DBACTYLASE SIRTUIN-TYPE.
FT ACT SITE 213 213 BY SIMILARITY.
FT METAL 221 221 ZINC (BY SIMILARITY).
FT METAL 224 224 ZINC (BY SIMILARITY).
FT METAL 251 251 ZINC (BY SIMILARITY).
FT METAL 254 254 ZINC (BY SIMILARITY).
SQ SEQUENCE 370 AA; 41785 MW; 86BB0238BFA914F1 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 370;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 KPISTV 224
DB 204 KPISTV 210

RESULT 29
YQ4S YEAST
ID YQ4S YEAST STANDARD; PRT; 409 AA.
AC P50082;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 46.7 kDa protein in PET54-DIE2 intergenic region.
GN YGR225W OR G841.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=98267763; PubMed=8701610;
RA van der Aart O.J.M., Klein K., Steensma H.Y.;
RT "Sequence analysis of the 43 kb CRM1-YLM9-PET54-DIE2-SM11-PHO81-YHB4-
RT PFK1 region from the right arm of Saccharomyces cerevisiae chromosome
RT VII."
RL Yeast 12:385-390 (1996).
CC -!- SIMILARITY: SOME, TO YEAST CDC20 AND S.POMBE SPAC1366.08.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X87941; CAA61174.1; -.
CC DR EMBL; Z73010; CAA97253.1; -.
CC DR PIR; S57689; S57689.
CC DR GerMOnline; 141537; -.
CC DR SGD; S0003457; AXA1.
CC DR InterPro; IPR000002; Fizzy.
CC DR ProDom; PD004563; Fizzy; 1.
KW Hypothetical protein.
SQ SEQUENCE 409 AA; 46660 MW; 91F7A246A28924D6 CRC64;

```

```

Query Match 2.9%; Score 7; DB 1; Length 409;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYS 94
DB 112 ASRNDYS 118

```

```

RESULT 30
YDML SCHPO
ID YDML SCHPO STANDARD; PRT; 441 AA.
AC O13909;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C23C11.01 in chromosome I.
GN SPAC23C11.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vansteens E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880 (2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: SOME, TO YEAST YIL090W.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98559; CAB11154.1; -.
CC DR PIR; T38239; T38239.
CC DR GenedB_Spombe; SPAC23C11.01; -.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 62 82 POTENTIAL.
CC FT TRANSMEM 88 108 POTENTIAL.
CC FT TRANSMEM 112 132 POTENTIAL.
CC FT TRANSMEM 154 174 POTENTIAL.
CC FT TRANSMEM 192 212 POTENTIAL.
CC FT TRANSMEM 224 244 POTENTIAL.
CC FT TRANSMEM 247 267 POTENTIAL.
CC FT TRANSMEM 312 332 POTENTIAL.
CC FT TRANSMEM 335 355 POTENTIAL.
CC FT TRANSMEM 363 383 POTENTIAL.
CC FT TRANSMEM 399 419 POTENTIAL.
CC SQ SEQUENCE 441 AA; 49169 MW; 12B9B0095A93B3AA CRC64;

```

Query Match

2.9%; Score 7; DB 1; Length 441;

Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TGOALAS 174  
DB 430 TGOALAS 436

RESULT 31  
AROA MYCTU  
ID AROA MYCTU STANDARD; PRT; 450 AA.  
AC P22457;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-  
enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
GN AROA OR RV3227 OR MT3324 OR MTCV20511.02.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=91072223; PubMed=2123856;  
RA Garbe T., Jones C., Charles I.G., Dougan G., Young D.;  
RT "Cloning and characterization of the *aroA* gene from *Mycobacterium*  
J. Bacteriol. 172:6774-6782(1990).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=9895987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaaia F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of *Mycobacterium tuberculosis* clinical and  
laboratory strains."  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =  
Phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.  
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
sixth step.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the EPSP synthase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X52269; CAA36510.1; -;  
DR EMBL; M62708; AAA25356.1; -;  
DR EMBL; Z95121; CAB0328.1; -;  
DR EMBL; AE007144; AAK47667.1; -;  
DR PIR; E70590; E70590.  
DR TIGR; MT3324; -;  
DR TubercuList; RV3227; -;  
DR HAVAP; MF 00210; -; 1.  
DR InterPro; IPR006264; AroA.  
DR InterPro; IPR001986; EPSP\_synth.  
DR Pfam; PF00275; EPSP\_synthase; 1.  
DR PRODOM; PD001867; EPSP\_synthase; 1.  
DR TIGRFAMs; TIGR01356; aroA; 1.  
DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.  
DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.  
KW Aromatic amino acid biosynthesis; transferase; Complete proteome.  
SQ SEQUENCE 450 AA; 46425 MW; 27B86F9412A07D5 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 450;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
DB 321 ALASPGS 327

RESULT 32  
CA44 BOVIN  
ID CA44 BOVIN STANDARD; PRT; 453 AA.  
AC Q29422;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Collagen alpha 4(IV) chain (Fragment).  
GN COL4A4.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.  
RC TISSUE=Lens;  
RA MEDLINE=92112769; PubMed=1370461;  
RA Mariyama M., Kalluri R., Hudson B.G., Readers S.T.;  
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of  
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV  
RT collagens."  
RL J. Biol. Chem. 267:1253-1258(1992).  
RN [2]  
RP SEQUENCE OF 217-246.  
RX MEDLINE=90202779; PubMed=2318822;  
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;  
RT "Glomerular basement membrane. Identification of a fourth chain,  
RT alpha 4, of type IV collagen."  
RL J. Biol. Chem. 265:5466-5469(1990).  
RN [3]  
RP SEQUENCE OF 217-233.  
RX MEDLINE=87222419; PubMed=2438283;  
RA Butkowski R.J., Langeveld J.P.M., Wierslander J., Hamilton J.,  
RA Hudson B.G.;  
RT "Localization of the Goodpasture epitope to a novel chain of basement  
RT membrane collagen."  
RL J. Biol. Chem. 262:7874-7877(1987).  
CC -!- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure with  
CC 2 other chains to generate type IV collagen network.  
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).  
CC -----





```

DR HSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
|||||

RESULT 37
HEMA_IADH3 STANDARD; PRT; 550 AA.
AC P12584; Q84012; Q89793;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16739; AAA43145.1; -
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC InterPro; IPR008975; Viral_cap_coat.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
KW Envelope protein; Hemagglutinin; Glycoprotein.

```

```

FT NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
|||||

RESULT 38
HEMA_IADH4 STANDARD; PRT; 550 AA.
AC P12585; Q84013; Q84014;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16740; AAA43146.1; -
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC InterPro; IPR008975; Viral_cap_coat.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
KW Envelope protein; Hemagglutinin; Glycoprotein.

```

```
SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364

RESULT 39
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5 Q84015; Q84016;
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawachi Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16741; AAA43147.1; -
DR PIR; E27813; HMIV21.
DR HSSP; P03437; 2VU.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15B9F8C CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;
Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364

RESULT 41
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7 Q84017; Q84017;
AC P12587; Q84017; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawachi Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16742; AAA43148.1; -
DR PIR; F27813; HMIV98.
DR HSSP; P03437; IHGJ.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
```

AC P12588; Q84018; Q89470;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain] (Fragment).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hokkaido/10/85).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OC NCBI\_TaxID=11363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87265458; PubMed=2440178;  
 RA Kida H., Kawakita Y., Naeve C.W., Webster R.G.;  
 RT "Antigenic and genetic conservation of H3 influenza virus in wild  
 RT ducks";  
 RL Virology 159:109-119(1987).  
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
 CC cell receptors and for initiating infection.  
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
 CC (HA1 and HA2) linked by a disulfide bond.  
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; M16743; AAA43149.1; -  
 DR HSSP; P03437; 3HWG.  
 DR InterPro; IPR008980; Capsid hemag.  
 DR InterPro; IPR001364; Hemagglutn.  
 DR Pfam; PF00509; Hemagglutinin; 1.  
 DR PRINTS; PR00329; Hemagglutinin; 1.  
 DR ProDom; PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON TER 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61761 MW; 65F81793281D53EB CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 165 SEGTGQA 171  
 DB 358 SEGTGQA 364  
 RESULT 42  
 HEMA\_IADHK  
 ID HEMA\_IADHK STANDARD; PRT; 550 AA.  
 AC P43257;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain] (Fragment).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.  
 OC NCBI\_TaxID=11364;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91341491; PubMed=1875195;  
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
 RT "Molecular evidence for a role of domestic ducks in the introduction  
 RT of avian H3 influenza viruses to pigs in southern China, where the  
 RT A/Hong Kong/68 (H3N2) strain emerged.";  
 RL J. Gen. Virol. 72:2007-2010(1991).  
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
 CC cell receptors and for initiating infection.  
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
 CC (HA1 and HA2) linked by a disulfide bond.  
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; D00929; BAA00769.1; -  
 DR HSSP; P03437; 2VIU.  
 DR InterPro; IPR008980; Capsid hemag.  
 DR InterPro; IPR001364; Hemagglutn.  
 DR Pfam; PF00509; Hemagglutinin; 1.  
 DR PRINTS; PR00329; Hemagglutn; 1.  
 DR ProDom; PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON TER 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61549 MW; 864639B829FE1B89 CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 165 SEGTGQA 171  
 DB 358 SEGTGQA 364  
 RESULT 43  
 HEMA\_IADHL  
 ID HEMA\_IADHL STANDARD; PRT; 550 AA.  
 AC P43258;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain] (Fragment).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hong Kong/64/76).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OC NCBI\_TaxID=45412;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91341491; PubMed=1875195;  
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;  
 RT "Molecular evidence for a role of domestic ducks in the introduction  
 RT of avian H3 influenza viruses to pigs in southern China, where the



```

RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00932; BAA00771.1; -.
DR HSSP; P03437; 2VU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; Hemagglutn12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 328 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61762 MW; 67EF8B4948C191A CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364
|||||

RESULT 45
HEMA_IAGHK
ID HEMA_IAGHK .STANDARD; PRT; 550 AA.
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45411;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00930; BAA00770.1; -.
DR HSSP; P03437; 2VU.

```

```
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00509; Hemagglutinin; I.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 46
HEMA_IAMB6 STANDARD; PRT; 550 AA.
AC P12589;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Memphis/6/86).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11440;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306236; PubMed=3407150;
RA Katz J.M., Webster R.G.;
RT "Antigenic and structural characterization of multiple subpopulations
of H3N2 influenza virus from an individual.";
RL Virology 165:446-456(1988).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
(HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M21648; AAA43275.1; -.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 358 SEGTGQA 364

RESULT 47
HEMA_IJZH2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/91/78).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11497;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China.";
RL Virology 162:160-166(1988).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
(HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M19057; AAA43212.1; -.
DR FIR; B29971; HMIVS3.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD00225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61804 MW; 52C9F14B309310ED CRC64;
```

```
SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364

RESULT 48
HEMA IAH3
ID HEMA IAH3 STANDARD; PRT; 550 AA.
AC P1134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
GN HA
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.";
RL Virology 162:160-166(1988).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19056; AAA43211.1; ALT_TERM.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Hemagglutinin; Envelope protein; Glycoprotein.
FT NON_TER 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 358 SEGTGQA 364

RESULT 50
HEMA IAHFO
ID HEMA IAHFO STANDARD; PRT; 565 AA.
AC P16995; Q83992; Q83993;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
GN HA
OS Influenza A virus (strain A/Equine/Algiers/72).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 163:283-292(1989).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24721; AAA43100.1; ALT_SEQ.
DR HSSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 340 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 373 SEGTGQA 379
```

```
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Fontainebleau/76) (Influenza A
OS virus (strain A/Equine/France/1/76)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11399;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138(1992)
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24723; AAA3101.1; ALT_SEQ.
DR EMBL; M73773; -; NOT_ANNOTATED_CDS.
DR HSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 13 15 WVY -> AVD (IN REF. 2).
FT CONFLICT 20 20 T -> I (IN REF. 2).
FT CONFLICT 150 150 R -> G (IN REF. 2).
FT CONFLICT 187 187 N -> D (IN REF. 2).
FT CONFLICT 242 242 S -> A (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 479 479 N -> G (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63686 MW; 1BB06B765992E87C CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTOQA 171
Db 373 SEGTOQA 379

RESULT 52
HEMA IAHK6
ID HEMA IAHK6 STANDARD; PRT; 565 AA.
AC P19699;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/2/86).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11403;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24727; -; NOT_ANNOTATED_CDS.
DR HSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 13 15 WVY -> AVD (IN REF. 2).
FT CONFLICT 20 20 T -> I (IN REF. 2).
FT CONFLICT 150 150 R -> G (IN REF. 2).
FT CONFLICT 187 187 N -> D (IN REF. 2).
FT CONFLICT 242 242 S -> A (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 479 479 N -> G (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63610 MW; 2038CC1C6C9B88C5 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEGTOQA 171
Db 373 SEGTOQA 379

RESULT 52
HEMA IAHK7
ID HEMA IAHK7 STANDARD; PRT; 565 AA.
AC P16956; Q83994; Q83995;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
```

```

GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/1/87).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24722; AAA43107.1; ALT_SEQ.
CC HSSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63702 MW; 93963AF456486787 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 53
HEMA_IAHNM STANDARD; PRT; 565 AA.
AC P16997; Q83996; Q83997;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Romania/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24722; AAA43107.1; ALT_SEQ.
CC HSSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63702 MW; 93963AF456486787 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 54
HEMA_IAHRO STANDARD; PRT; 565 AA.
AC P16998; Q83998; Q83999;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Romania/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M24724; AAA43109.1; ALT\_SEQ.  
 DR HSSP; P03437; 1HTM.  
 DR InterPro; IPR008980; Capsid hemag.  
 DR InterPro; IPR001364; Hemagglutn.  
 DR Pfam; PF00509; Hemagglutinin; 1.  
 DR PRINTS; PR00329; HEMAGGLUTN12.  
 DR ProDom; PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 565 AA; 63660 MW; 91A07D33FAFDC842 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 SEGTGQA 171  
 DB 373 SEGTGQA 379

## RESULT 55

HEMA\_IAHSU  
 ID HEMA\_IAHSU STANDARD; PRT; 565 AA.  
 AC P16999; Q84000; Q84001;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain].  
 GN HA.  
 OS Influenza A virus (strain A/Equine/Santiago/1/85).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=11414;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89204899; PubMed=2705299;  
 RA Kawaoka Y., Bean W.J., Webster R.G.;  
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";  
 RL Virology 169:283-292(1989).  
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
 CC cell receptors and for initiating infection.  
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
 CC (HA1 and HA2) linked by a disulfide bond.  
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M24725; AAA43110.1; ALT\_SEQ.  
 DR HSSP; P03437; 1HTM.  
 DR InterPro; IPR008980; Capsid hemag.  
 DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.  
 DR PRINTS; PR00329; HEMAGGLUTN12.  
 DR ProDom; PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 565 AA; 63665 MW; 399F4BF4BA231327 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 SEGTGQA 171  
 DB 373 SEGTGQA 379

## RESULT 56

HEMA\_IAHSU  
 ID HEMA\_IAHSU STANDARD; PRT; 565 AA.  
 AC Q08011;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
 DE Hemagglutinin HA2 chain].  
 GN HA.  
 OS Influenza A virus (strain A/Equine/Suffolk/89).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=45413;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93277383; PubMed=8503788;  
 RA Birns M.M., Daly J.M., Chirnside E.D., Mumford J.A., Wood J.M.,  
 RA Richards C.M., Daniels R.S.;  
 RT "Genetic and antigenic analysis of an equine influenza H 3 isolate  
 RT from the 1989 epidemic.";  
 RL Arch. Virol. 130:33-44(1993).  
 CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
 CC cell receptors and for initiating infection.  
 CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
 CC (HA1 and HA2) linked by a disulfide bond.  
 CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X58437; CAA48482.1; --  
 DR HSSP; P03437; 1HTM.  
 DR InterPro; IPR008980; Capsid hemag.  
 DR InterPro; IPR001364; Hemagglutn.  
 DR Pfam; PF00509; Hemagglutinin; 1.  
 DR PRINTS; PR00329; HEMAGGLUTN12.  
 DR ProDom; PD000225; Hemagglutn; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN (BY SIMILARITY).  
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN (BY SIMILARITY).  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63699 MW; C7A4E3B54B87D1A1 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 57
HEMA IAHTO
ID HEMA IAHTO STANDARD; PRT; 565 AA.
AC P17001; Q84004; Q84005;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tennessee/5/86)
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11417;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawachi Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC HSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63698 MW; 1F54485F0E7AC2C4 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379
```

```
Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379

RESULT 58
HEMA IAHTO
ID HEMA IAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawachi Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24720; AAA43111.1; ALT_SEQ.
CC HSP; P03437; IHTM.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63580 MW; 84BD7AD7062937A CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379
```



```

RESULT 59
HEMA_IAHUR STANDARD; PRT; 565 AA.
AC P17002; Q84006; Q84007;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Equine/Uruguay/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=27052299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; M24718; AAA43114.1; ALT_SEQ.
DR HSP; P03437; IHTM.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL
FT CHAIN
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63604 MW; 32818573564C8F94 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171
Db 373 SEGTGQA 379
|||||

RESULT 60
HEMA_IATAC STANDARD; PRT; 566 AA.
AC P03437;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).

```

```

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
RA Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
RT influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL Nature 286:771-776(1980).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=81123029; PubMed=7464906;
RA Wilson I.A., Skehel J.J., Wiley D.C.;
RT "Structure of the haemagglutinin membrane glycoprotein of influenza
RT virus at 3-A resolution.";
RL Nature 289:366-373(1981).
RN [3]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=88232903; PubMed=3374584;
RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA Wiley D.C.;
RT "Structure of the influenza virus haemagglutinin complexed with its
RT receptor, sialic acid.";
RL Nature 333:426-431(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX MEDLINE=90107940; PubMed=2295311;
RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA Wiley D.C.;
RT "The structure of a membrane fusion mutant of the influenza virus
RT haemagglutinin.";
RL EMBO J. 9:17-24(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=90230310; PubMed=2329580;
RA Weis W.I., Bruenger A.T., Skehel J.J., Wiley D.C.;
RT "Refinement of the influenza virus hemagglutinin by simulated
RT annealing.";
RL J. Mol. Biol. 212:737-761(1990).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94352388; PubMed=8072525;
RA Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL Nature 371:37-43(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3-25 ANGSTROMS).
RX MEDLINE=98120975; PubMed=9461077;
RA Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT "Antigen distortion allows influenza virus to escape neutralization.";
RL Nat. Struct. Biol. 5:119-123(1998).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; J02090; AAA43178.1; -.
DR EMBL; V01085; CAA24269.1; -.
DR PIR; A33231; HMVHA.
DR PDB; 2HMG; 31-OCT-93.
DR PDB; 3HMG; 31-OCT-93.
DR PDB; 4HMG; 31-OCT-93.

```

DR PDB; SHMG; 31-JAN-94.  
DR PDB; IHGD; 31-JAN-94.  
DR PDB; IHGE; 31-JAN-94.  
DR PDB; IHGF; 31-JAN-94.  
DR PDB; IHGG; 31-JAN-94.  
DR PDB; IHGH; 31-JAN-94.  
DR PDB; IHGI; 31-JAN-94.  
DR PDB; IHGJ; 31-JAN-94.  
DR PDB; IHM; 14-FEB-95.  
DR PDB; IHM; 29-APR-98.  
DR PDB; 2VIS; 29-APR-98.  
DR PDB; 2VIT; 29-APR-98.  
DR PDB; 2VIU; 29-APR-98.  
DR PDB; 1EO8; 29-NOV-00.  
DR PDB; 1HA0; 22-DEC-99.  
DR PDB; 1JBH; 13-MAR-02.  
DR PDB; 1KEN; 24-APR-02.  
DR PDB; 1QFU; 29-DEC-99.  
DR PDB; 1QUL; 05-JAN-00.  
DR InterPro; IPR008980; Capsid hemag.  
DR Pfam; PF00509; Hemagglutinin: 1.  
DR PRINTS; PR00329; HEMAGGLUTN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.  
FT SIGNAL 1 16  
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.  
FT DISULFID 30 482 INTERCHAIN.  
FT DISULFID 68 293  
FT DISULFID 80 92  
FT DISULFID 113 155  
FT DISULFID 297 321  
FT DISULFID 489 493  
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .).  
FT STRAND 27 35  
FT STRAND 40 42  
FT STRAND 50 52  
FT STRAND 55 57  
FT STRAND 59 60  
FT STRAND 67 70  
FT STRAND 74 76  
FT TURN 78 79  
FT TURN 82 87  
FT TURN 88 88  
FT HELIX 90 95  
FT TURN 96 97  
FT TURN 99 99  
FT STRAND 102 105  
FT TURN 107 108  
FT STRAND 116 117  
FT TURN 119 120  
FT HELIX 121 131  
FT STRAND 133 133  
FT STRAND 136 138  
FT TURN 144 145  
FT STRAND 146 147  
FT STRAND 152 157  
FT TURN 158 159  
FT STRAND 160 162  
FT TURN 165 166  
FT STRAND 167 169  
FT STRAND 171 173  
FT TURN 174 175  
FT STRAND 176 176  
FT STRAND 180 185  
FT STRAND 192 200

204 211  
217 221  
226 229  
239 239  
240 241  
242 242  
245 253  
255 256  
258 265  
267 270  
272 275  
282 285  
290 294  
297 299  
300 301  
302 304  
310 311  
318 320  
323 324  
331 333  
337 337  
347 348  
350 350  
351 351  
352 354  
355 355  
359 360  
361 361  
367 373  
374 375  
376 382  
383 400  
401 401  
406 407  
421 471  
421 471  
472 474  
475 477  
482 485  
491 498  
499 500  
504 515  
518 519  
566 AA; 63415 MW; E395659C23CAFECA CRC64;  
SQ SEQUENCE

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 374 SEGTGQA 380

RESULT 61  
HEMA\_IADA3 STANDARD; PRT; 566 AA.  
AC P26134;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain].  
GN HA  
OS Influenza A virus (strain A/Duck/Alberta/78/76).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=11348;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92114135; PubMed=1731092;  
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,  
RA Webster R.G.;  
RT "Evolution of the H3 influenza virus hemagglutinin from human and

RT nonhuman hosts.";  
RL J. Virol. 66:1129-1138 (1992).  
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
CC cell receptors and for initiating infection.  
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
CC (HA1 and HA2) linked by a disulfide bond.  
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M73771; -; NOT\_ANNOTATED\_CDS.  
DR HSSP; P03437; 2VIU.  
DR InterPro; IPR008980; Capsid hemag.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 374 SEGTGQA 380  
RESULT 62  
HEMA\_IADM2  
ID HEMA\_IADM2 STANDARD; PRT; 566 AA.  
AC P26135;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain].  
GN HA.  
OS Influenza A virus (strain A/Duck/Memphis/328/74).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=11367;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92114135; PubMed=1731092;  
RA Bean W.J., Schell M., Katz J., Kawasaka Y., Naeve C., Gorman O.,  
RA Webster R.G.;  
RT "Evolution of the H3 influenza virus hemagglutinin from human and  
RT nonhuman hosts."  
RL J. Virol. 66:1129-1138 (1992).  
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
CC cell receptors and for initiating infection.  
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
CC (HA1 and HA2) linked by a disulfide bond.  
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M73772; -; NOT\_ANNOTATED\_CDS.  
DR HSSP; P03437; 2VIU.  
DR InterPro; IPR008980; Capsid hemag.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63572 MW; C85DFCSDESBD80D CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 165 SEGTGQA 171  
DB 374 SEGTGQA 380  
RESULT 63  
HEMA\_IADU3  
ID HEMA\_IADU3 STANDARD; PRT; 566 AA.  
AC P03442;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain].  
GN HA.  
OS Influenza A virus (strain A/Duck/Ukraine/1/63).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=11374;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82025542; PubMed=6169439;  
RA Fang R., Min Jou W., Huybrecock D., Devos R., Fiers W.;  
RA "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin  
RT gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza  
RT hemagglutinin."  
RL Cell 25:315-323 (1981).  
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
CC cell receptors and for initiating infection.  
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
CC (HA1 and HA2) linked by a disulfide bond.  
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

```
-----
EMBL; V01087; CAA24271.1; -.
PDB; 11BN; 08-AUG-01.
PDB; 11BO; 08-AUG-01.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00509; Hemagglutinin; I.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 374 SEGTGOA 380
|||||

RESULT 64
HEMA IAEN7
ID HEMA IAEN7 STANDARD; PRT; 566 AA.
AC P03440;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/England/321/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11378;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83110955; PubMed=6822816;
RA Hauptmann R., Clarke L.D., Mountford R.C., Bachmayer H., Almond J.W.;
RT "Nucleotide sequence of the haemagglutinin gene of influenza virus A/England/321/77.";
RL J. Gen. Virol. 64:215-220 (1983).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
-----
EMBL; X05907; CAA29337.1; -.
DR PIR; B92790; HMTV6.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.

```

```
PRINTS; PR00329; HEMAGGLUTN12.
ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63608 MW; FA5B886FF4B8C888 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171
DB 374 SEGTGOA 380
|||||

RESULT 65
HEMA IAMAO
ID HEMA IAMAO STANDARD; PRT; 566 AA.
AC P26138;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain; Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Mallard/New York/6874/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11436;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O., Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and nonhuman hosts.";
RL J. Virol. 66:1129-1138 (1992).
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to cell receptors and for initiating infection.
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains (HA1 and HA2) linked by a disulfide bond.
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
-----
EMBL; M73776; -. NOT_ANNOTATED_CDS.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).

```

```
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63574 MW; BB206011COBD9A3B CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 374 SEGTGQA 380

RESULT 66
HEMA_IAME1 STANDARD; PRT; 566 AA.
AC P03439;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Memphis/1/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11438;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83303827; PubMed=6193634;
RA Newton S.E., Air G.M., Webster R.G., Laver W.G.;
RT "Sequence of the hemagglutinin gene of influenza virus A/Memphis/1/71
RT and previously uncharacterized monoclonal antibody-derived
RT variants."
RL Virology 128:495-501 (1983).
RN [2]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=82150925; PubMed=6174976;
RA Air G.M.;
RT "Sequence relationships among the hemagglutinin genes of 12 subtypes
RT of influenza A virus."
RL Proc. Natl. Acad. Sci. U.S.A. 78:7639-7643 (1981).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02132; AAA43187.1; -.
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; Hemagglutinin; 1.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 346 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63264 MW; 1D9313AB3C380CD7 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
```

```
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63558 MW; 6383C6F9D81AB941 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 374 SEGTGQA 380

RESULT 67
HEMA_IAME2 STANDARD; PRT; 566 AA.
AC P03439;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Memphis/102/72).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11439;
RN [1]
RP SEQUENCE FROM N.A.
RA Sleight M.J., Both G.W., Brownlee G.G., Bender V.J., Moss B.A.;
RT "The haemagglutinin gene of influenza A virus: nucleotide sequence
RT analysis of cloned DNA copies."
RL (In) Laver G., Air G. (eds.);
RL Structure and variation in influenza virus, pp.69-79, Elsevier,
RL New York (1980).
CC -1- FUNCTION: Hemagglutinin is responsible for attaching the virus to
CC cell receptors and for initiating infection.
CC -1- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains
CC (HA1 and HA2) linked by a disulfide bond.
CC -1- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01089; CAA24273.1; -.
CC FIR; A94441; HMTVHM.
CC HSP; P03437; 2VIU.
CC InterPro; IPR008980; Capsid hemag.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; Hemagglutn; 1.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC SIGNAL 1 16
CC CHAIN 17 346 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63264 MW; 1D9313AB3C380CD7 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;
```



Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 374 SEGTGQA 380

RESULT 70  
HEMA IAZUK STANDARD; PRT; 566 AA.  
AC P26141;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain].  
GN HA.

OS Influenza A virus (strain A/Swine/Ukkel/1/84).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=11517;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=92114135; PubMed=1731092;  
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,  
RA Webster R.G.;  
RT "Evolution of the H3 influenza virus hemagglutinin from human and  
RT nonhuman hosts";  
RL J. Virol. 66:1129-1138 (1992).  
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
CC cell receptors and for initiating infection.  
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
CC (HA1 and HA2) linked by a disulfide bond.  
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M73775; -; NOT ANNOTATED\_CDS.  
CC HSSP; P03437; 1HTM.  
CC InterPro; IPR008980; Capsid hemag.  
CC InterPro; IPR001364; Hemagglutn.  
CC Pfam; PF00509; Hemagglutinin; 1.  
CC PRINTS; PR00329; HEMAGGLUTN12.  
CC ProDom; PD000225; Hemagglutn; 1.  
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
CC SIGNAL 1 16  
CC CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.  
CC FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.  
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 38 28 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SEQUENCE 566 AA; 63725 MW; 44661ED8B3D5B331 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 566;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 374 SEGTGQA 380

RESULT 71  
HEMA IAVI7 STANDARD; PRT; 567 AA.  
AC P03435;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
DE Hemagglutinin HA2 chain].  
GN HA.  
OS Influenza A virus (strain A/Victoria/3/75).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=11483;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=80155186; PubMed=6153930;  
RA Min Jou W., Verhoeven M., Devos R., Saman E., Fang R.,  
RA Huybrecock D., Fiers W., Threlfall G., Barber C., Carey N.,  
RA Entage S.;  
RT "Complete structure of the hemagglutinin gene from the human  
RT Influenza A/Victoria/3/75 (H3N2) strain as determined from Cloned  
RT DNA";  
RL Cell 19:683-696 (1980).  
RN [2]  
SEQUENCE FROM N.A.  
RX MEDLINE=80354693; PubMed=7402351;  
RA Verhoeven M., Fang R., Jou W.M., Devos R., Huybrecock D., Saman E.,  
RA Fiers W.;  
RT "Antigenic drift between the haemagglutinin of the Hong Kong  
RT influenza strains A/Aichi/2/68 and A/Victoria/3/75";  
RL Nature 286:771-776 (1980).  
CC -!- FUNCTION: Hemagglutinin is responsible for attaching the virus to  
CC cell receptors and for initiating infection.  
CC -!- SUBUNIT: Homotrimer. Each of the monomers is formed by two chains  
CC (HA1 and HA2) linked by a disulfide bond.  
CC -!- SIMILARITY: Belongs to the influenza viruses hemagglutinin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; V01098; CAA24281.1; -  
CC EMBL; V01085; CAA24270.1; -  
CC PIR; A90794; HMIV.  
CC HSSP; P03437; 2VIU.  
CC InterPro; IPR008980; Capsid hemag.  
CC InterPro; IPR001364; Hemagglutn.  
CC Pfam; PF00509; Hemagglutinin; 1.  
CC PRINTS; PR00329; HEMAGGLUTN12.  
CC ProDom; PD000225; Hemagglutn; 1.  
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
CC SIGNAL 1 16  
CC CHAIN 17 345 HEMAGGLUTININ HA1 CHAIN.  
CC FT CHAIN 347 567 HEMAGGLUTININ HA2 CHAIN.  
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 25 25 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 55 55 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 182 182 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SEQUENCE 567 AA; 63422 MW; 824D98A880EC5DEF CRC64;

Query Match 2.9%; Score 7; DB 1; Length 567;



```
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTQQA 171
    |||||
Db 375 SEGTQQA 381

RESULT 72
CA44 RABIT
ID CA44 RABIT STANDARD; PRT; 623 AA.
AC P55787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
TX TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01477; -; NOT_ANNOTATED_CDS.
CC
CC PIR; A45137; A45137.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001442; Procollagen4_C.
CC Pfam; PF01413; C4; 2.
CC Pfam; PF01391; Collagen; 5.
CC ProDom; PD003923; ProcollagenC4; 1.
CC SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT DOMAIN 393 623 NONHELICAL REGION (NC1).
FT DISULFID 413 502 OR 499 (BY SIMILARITY).
```

```
FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 484 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SQ SEQUENCE 623 AA; 62393 MW; CCBC9BB31242FE82 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 623;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
    |||||
Db 551 SPGSCLE 557

RESULT 73
TOXA_PSEAE
ID TOXA_PSEAE STANDARD; PRT; 638 AA.
AC P11439; Q91417;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exotoxin A precursor (NAD-dependent ADP-ribosyltransferase)
DE (EC 2.4.2.-).
GN ETA OR PA1148.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 26-53.
RX MEDLINE=84194063; PubMed=6201861;
RA Gray G.L., Smith D.H., Baldridge J.S., Harkins R.N., Vasil M.L.,
RA Chen E.Y., Heyneker H.L.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT the exotoxin A structural gene of Pseudomonas aeruginosa.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:2645-2649(1984).
RN [2]
SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Stover C.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sauer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [3]
ACTIVE SITE.
RX MEDLINE=87250491; PubMed=2885323;
RA Carroll S.F., Collier R.J.;
RT "Active site of Pseudomonas aeruginosa exotoxin A. Glutamic acid 553
RT is photolabeled by NAD and shows functional homology with glutamic
RT acid 148 of diphtheria toxin.";
RN J. Biol. Chem. 262:8707-8711(1987).
RN [4]
DOMAINS.
RX MEDLINE=90375493; PubMed=2118903;
RA Chaudhary V.K., Jinno Y., Galo M.G., Fitzgerald D., Pastan I.;
RT "Mutagenesis of Pseudomonas exotoxin in identification of sequences
RT responsible for the animal toxicity.";
RN J. Biol. Chem. 265:16306-16310(1990).
RN [5]
DOMAINS.
RX MEDLINE=91006124; PubMed=2170123;
RA Bourdenet S., Vacheron M.-J., Guinand M., Michel G., Arminjon F.;
RT "Biochemical and immunochemical studies of proteolytic fragments of
RT exotoxin A from Pseudomonas aeruginosa.";
RL Eur. J. Biochem. 192:379-385(1990).
```

RN [6] DISULFIDE BOND.  
RP MEDLINE=20068844; PubMed=10600112;  
RA McKee M.L., FitzGerald D.J.;  
RT "Reduction of furin-nicked Pseudomonas exotoxin A: an unfolding  
RT story.";  
RL Biochemistry 39:16507-16513(1999).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 424-638.  
RX MEDLINE=96016159; PubMed=7568123;  
RA Li M., Dyda F., Benhar I., Pastan I., Davies D.R.;  
RT "The crystal structure of Pseudomonas aeruginosa exotoxin domain III  
RT with nicotinamide and AMP: conformational differences with the intact  
RT exotoxin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:9308-9312(1995).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 424-638.  
RX MEDLINE=96293446; PubMed=8692916;  
RA Li M., Dyda F., Benhar I., Pastan I., Davies D.R.;  
RT "Crystal structure of the catalytic domain of Pseudomonas exotoxin A  
RT complexed with a nicotinamide adenine dinucleotide analog:  
RT implications for the activation process and for ADP ribosylation.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:6902-6906(1996).  
CC -!- FUNCTION: THIS TOXIN IS AN NAD-DEPENDENT ADP-RIBOSYLTRANSFERASE.  
CC IT CATALYZES THE TRANSFER OF THE ADP RIBOSYL MOIETY OF OXIDIZED  
CC NAD ONTO ELONGATION FACTOR 2 (EF-2) THUS ARRESTING PROTEIN  
CC SYNTHESIS.  
CC -!- PTM: THE 8 CYSTEINES PARTICIPATE IN INTRACHAIN DISULFIDE BONDS.  
CC -!- SIMILARITY: REGIONAL SEQUENCE SIMILARITY AT THE ACTIVE SITE  
CC WITH DIPHTHERIA TOXIN (DT).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; K01397; AAB59097.1; -;  
DR EMBL; AE004544; AG04537.1; -;  
DR PIR; A30347; A30347.  
DR PIR; C83503; C83503.  
DR PDB; 1AER; 10-JUN-96.  
DR PDB; 1DNA; 15-SEP-95.  
DR InterPro; IPR008985; ConA\_like\_lec\_g1.  
DR PDB; 1IKQ; 12-DEC-01.  
DR PDB; 1IKQ; 12-DEC-01.  
KW Toxin; Signal; Transferase; Glycosyltransferase; NAD; 3D-structure;  
KW Complete proteome.  
FT SIGNAL 1 25  
FT CHAIN 26 638  
FT DOMAIN 26 277  
FT  
FT DOMAIN 278 389  
FT  
FT DOMAIN 390 429  
FT DOMAIN 430 638  
FT ACT\_SITE 465 465  
FT ACT\_SITE 578 578  
FT DISULFID 290 312  
FT CONFLICT 4 4  
FT CONFLICT 22 22  
FT CONFLICT 204 204  
FT CONFLICT 389 389  
FT CONFLICT 432 432  
FT CONFLICT 540 540  
FT STRAND 433 433  
FT TURN 436 437  
FT STRAND 440 440  
FT TURN 441 441  
FT HELIX 444 456  
FT TURN 457 458

FT STRAND 459 467  
FT HELIX 469 476  
FT TURN 477 478  
FT TURN 490 491  
FT STRAND 494 497  
FT HELIX 500 504  
FT TURN 505 506  
FT STRAND 508 508  
FT TURN 514 515  
FT STRAND 520 520  
FT STRAND 522 529  
FT HELIX 530 535  
FT STRAND 536 538  
FT TURN 543 544  
FT TURN 546 547  
FT HELIX 548 556  
FT TURN 557 557  
FT STRAND 566 570  
FT TURN 572 574  
FT STRAND 577 581  
FT TURN 583 586  
FT TURN 587 588  
FT STRAND 590 593  
FT TURN 600 601  
FT TURN 603 604  
FT HELIX 609 611  
FT HELIX 614 617  
FT TURN 618 619  
FT STRAND 626 626  
SQ SEQUENCE 638 AA; 69284 MW; 7B9AAD56A27C700A CRC64;  
  
Query Match 2.9%; Score 7; DB 1; Length 638;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 171 ALASPGS 177  
DB 340 ALASPGS 346  
  
RESULT 74  
SPH2 HUMAN  
ID SPH2 HUMAN STANDARD; PRT; 654 AA.  
AC Q9NRAQ; Q9BEN1; Q9HQQ2; Q9NWU7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Spingosine kinase 2 (EC 2.7.1.1-) (SK 2) (SPK 2).  
GN SPHK2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.  
RX MEDLINE=20347850; PubMed=10751414;  
RA Liu H., Sugura M., Nava V.E., Edsall L.C., Kono K., Foulton S.,  
RA Milstien S., Kohama T., Spiegel S.;  
RT "Molecular cloning and functional characterization of a novel  
RT mammalian sphingosine kinase type 2 isoform.";  
RL J. Biol. Chem. 275:19513-19520(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RT TISSUE=Brain;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
RA Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RT "Towards a catalog of human genes and proteins: sequencing and  
RT analysis of 500 novel complete protein coding human cDNAs.";  
RL Genome Res. 11:422-435(2001).

DR GO: 0008283; P: cell proliferation; NAS.  
DR GO: 0006869; P: sphinganine-1-phosphate biosynthesis; NAS.  
DR InterPro; IPR001206; DAGRC.  
DR InterPro; IPR007110; Ig-like.  
DR Pfam; PF00781; DAGKC; 1.  
DR ProDom; PD005043; DAGKC; 1.  
DR SMART; SM00046; DAGKC; 1.  
DR Transferase; Kinase; ATP-binding; Alternative splicing.  
FT VARSPLIC 1 36 Missing (in isoform 2 and isoform 3).  
FT VARSPLIC 292 390 /FTIDVSP 006217.  
FT LSVWGFVSDVDIQSERFRALASARFTLGLATLHYR  
FT LSVWGFVSDVDIQSERFRALASARFTLGLATLHYR  
FT GRUSYLPATVEPASPT -> PREDSDSTSSACPLWTTA  
FT RSCPERAASMPGSCPLLPOOLALGFRFIODRUNGGGRIG  
FT SLTCRGHTQRTLPAPAREGGSLFLKNINVFICKKKK  
FT (in isoform 3).  
FT /FTIDVSP 006218.  
FT P -> S (IN REF. 2).  
FT CONFLICT 49 49  
SQ SEQUENCE 654 AA; 69217 MW; F73PFC930DA50F CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 654;  
Best Local Similarity 100.0%; Pred. No. 44; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
Qy 171 ALASPGS 177  
Db 431 ALASPGS 437  
RESULT 75  
ID IGAA YERPE STANDARD; PRT; 715 AA.  
AC P58722;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative membrane protein igaa homolog.  
GN YPO0142 OR Y3922.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIM5 / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (Potential).  
CC -!- SIMILARITY: BELONGS TO THE IGAA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; AJ414141; CAC89005.1; -;  
 DR EMBL; AE013997; AAM87466.1; -;  
 DR PIR; AC0018; AC0018.  
 KW Transmembrane; Inner membrane; Complete proteome.  
 FT TRANSMEM 2 22 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 235 255 POTENTIAL.  
 FT TRANSMEM 349 369 POTENTIAL.  
 FT TRANSMEM 663 683 POTENTIAL.  
 SQ SEQUENCE 715 AA; 80005 MW; 5A2CB380FD1049F4 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 715;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TRGFVFT 23

DB 422 TRGFVFT 428

# RESULT 76

CA44 HUMAN

ID CA44 HUMAN STANDARD; PRT; 1690 AA.

AC P53420;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Collagen alpha 4(IV) chain precursor.

GN COL4A4.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=95014445; PubMed=7523402;

RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reeders S.T.;

RT "Complete primary structure of the human type IV collagen alpha 4(IV)

RT chain. Comparison with structure and expression of the other alpha

RT (IV) chains.";

RL J. Biol. Chem. 269:26172-26177 (1994).

RN [2]

RP SEQUENCE OF 1-23 FROM N.A.

RX MEDLINE=98196854; PubMed=9537506;

RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,

RA Ninomiya Y.;

RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and

RT alpha4(IV) collagen chains are arranged head-to-head on chromosome

RT 2q36.";

RL FEBS Lett. 424:11-16 (1998).

RN [3]

RP SEQUENCE OF 1219-1690 FROM N.A.

RX TISSUE=Eye;

RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;

RT "cDNA isolation and partial gene structure of the human alpha 4(IV)

RT collagen chain.";

RL FEBS Lett. 330:122-128 (1993).

RN [4]

RP SEQUENCE OF 1407-1507 FROM N.A.

RX MEDLINE=93054733; PubMed=1429714;

RA Kamagata Y., Mattei M.-G., Ninomiya Y.;

RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the

RT alpha 4 chain of basement membrane collagen type IV and assignment of

RT the gene to the distal long arm of human chromosome 2.";

RL J. Biol. Chem. 267:23753-23758 (1992).

RN [5]

RP

REVIEW ON VARIANTS.

RX MEDLINE=97338662; PubMed=9195222;

RA Lemmink H.H., Schroeder C.H., Momms L.A.H., Smeets H.J.M.;

RT "The clinical spectrum of type IV collagen mutations.";

RL Hum. Mutat. 9:477-493 (1997).

RN [6]

RN VARIANT AS SER-1201.

RP MEDLINE=95078927; PubMed=7987396;

RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,

RA Pirsion Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,

RA Smeets H.J.M., Reenders S.T.;

RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)

RT collagen genes in autosomal recessive Alport syndrome.";

RL Nat. Genet. 8:77-82 (1994).

RN [7]

RN VARIANT FBH GLU-897.

RP MEDLINE=96379660; PubMed=8787673;

RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,

RA Brunner H.G., van Oost B.A., Momms L.A.H., Smeets H.J.M.;

RT "Benign familial hematuria due to mutation of the type IV collagen

RT alpha4 gene.";

RL J. Clin. Invest. 98:1114-1118 (1996).

RN [8]

RN VARIANTS AS, AND VARIANTS.

RP MEDLINE=99011253; PubMed=9792860;

RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,

RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;

RT "Determination of the genomic structure of the COL4A4 gene and of

RT novel mutations causing autosomal recessive Alport syndrome.";

RL Am. J. Hum. Genet. 63:1329-1340 (1998).

CC -1- FUNCTION: Type IV collagen is the major structural component of

CC glomerular basement membranes (GBM), forming a 'chicken-wire'

CC meshwork together with laminins, proteoglycans and entactin/

CC nidogen.

CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-

CC alpha 6(IV), each of which can form a triple helix structure with

CC 2 other chains to generate type IV collagen network.

CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).

CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are

CC colocalized and present only in basement membranes of kidney, eye,

CC cochlea, lung and brain.

CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous

CC domain (NC1) at their C-terminus, frequent interruptions of the G-

CC X-Y repeats in the long central triple-helical domain (which may

CC cause flexibility in the triple helix), and a short N-terminal

CC triple-helical 7S domain.

CC -1- PTM: Prolines at the third position of the tripeptide repeating

CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -1- PTM: Type IV collagens contain numerous cysteine residues which

CC are involved in inter- and intramolecular disulfide bonding. 12 of

CC these, located in the NC1 domain, are conserved in all known type

CC IV collagens.

CC -1- DISEASE: Defects in COL4A4 are a cause of autosomal recessive

CC Alport syndrome (AS) [MIM:203780], an hereditary disorder

CC characterized by progressive glomerulonephritis, renal failure,

CC hematuria, ocular abnormalities and deafness. The recessive form

CC occurs equally between males and females.

CC -1- DISEASE: Defects in COL4A4 are a cause of familial benign

CC hematuria (FBH) [MIM:141200] or thin basement membrane disease.

CC FBH is characterized by persistent hematuria, an electron

CC microscopically detectable thin glomerular basement membrane (GBM)

CC and an autosomal dominant mode of inheritance. Renal function

CC remains normal. In children, differentiation between FBH and AS

CC can be difficult, because both disorders are manifested by

CC persistent hematuria and thin GBM at that age.

CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

```
CC CC
DR EMBL; X81053; CAA56943.1; -.
DR EMBL; AB008496; BAA25065.1; -.
DR EMBL; D17391; BAA04214.1; -.
DR PIR; A55360; CGHUIB.
DR Genew; HGNC:2206; COL4A4.
DR MIM; 120131; -.
DR MIM; 143200; -.
DR MIM; 203780; -.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38
FT CHAIN 39 1690
FT DOMAIN 39 64
FT DOMAIN 65 1459
FT DOMAIN 1460 1690
FT SITE 94 96
FT SITE 145 147
FT SITE 189 191
FT SITE 310 312
FT SITE 724 726
FT SITE 785 787
FT SITE 989 991
FT SITE 1206 1207
FT SITE 1212 1214
FT SITE 1480 1569
FT DISULFID 1513 1566
FT DISULFID 1525 1531
FT DISULFID 1588 1686
FT DISULFID 1622 1683
FT DISULFID 1634 1641
FT CARBOHYD 142 142
FT CARBOHYD 669 669
FT VARIANT 441 446
FT VARIANT 545 545
FT VARIANT 570 570
FT VARIANT 897 897
FT VARIANT 931 931
FT VARIANT 1004 1004
FT VARIANT 1030 1030
FT VARIANT 1201 1201
FT VARIANT 1402 1402
FT VARIANT 1572 1572
FT CONFLICT 1659 1660
FT SEQUENCE 1690 AA; 164095 MW; ELE72F283A72BAE CRC64;
Query Match 2.9%; Score 7; DB 1; Length 1690;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 174 SPSCLE 180
|||||
```

Db 1618 SPSCLE 1624

#### RESULT 77

```
CA24 CAEEL STANDARD; PRT; 1758 AA.
ID P17140; Q19098; Q19099;
AC 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (lethal protein 2).
GN LET-2 OR CLB-1 OR FO1G12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guc X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Wu X., Le T.T.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMBO J. 13:3278-3285(1994).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
CC intermolecular interactions between 7S domains and between NCI
CC domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I; Synonyms=a;
CC IsoId=P17140-1; Sequence=Displayed;
CC Name=II; Synonyms=b;
CC IsoId=P17140-2; Sequence=VSP_001160;
CC -!- DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NCI) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NCI domain, are conserved in all known type
CC IV collagens.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

EMBL; Z22964; CAA80536.1; -	48	VARIANT	48	G -> E (IN MN114; 73% LETHAL).
EMBL; Z22964; CAA80537.1; -	366	VARIANT	366	A -> T (IN MN126; 100% LETHAL).
EMBL; J05066; AAA27989.1; -	570	VARIANT	570	G -> E (IN MN109; 37% LETHAL).
EMBL; U22327; AAA64312.1; ALT_SEQ.	588	VARIANT	588	G -> R (IN MN103 AND MN151; 96% LETHAL).
EMBL; U53342; AAA96215.1; -	597	VARIANT	597	G -> R (IN MN132; 50% LETHAL).
EMBL; U53342; AAA96216.1; -	690	VARIANT	690	G -> R (IN MN101; 100% LETHAL).
PIR; T29350; T29350.	690	VARIANT	690	G -> E (IN MN129; 100% LETHAL).
WormPep; F01G12.5a; CE04334.	737	VARIANT	737	G -> E (IN MN143; 100% LETHAL).
WormPep; F01G12.5b; CE04335.	877	VARIANT	877	G -> R (IN G30; 90% LETHAL).
GO; GO:0005587; C:collagen type IV; IMP.	904	VARIANT	904	G -> R (IN E1470; 94% LETHAL).
GO; GO:0030203; F:extracellular matrix structural constituent. .; IMP.	1003	VARIANT	1003	G -> E (IN MN139; 20% LETHAL).
GO; GO:0016043; P:cell organization and biogenesis; NAS.	1125	VARIANT	1125	G -> D (IN G25; 2% LETHAL).
InterPro; IPR008161; Clg_helix.	1152	VARIANT	1152	G -> D (IN G17 AND B246; 9% LETHAL).
InterPro; IPR008160; Collagen.	1286	VARIANT	1286	G -> D (IN REF. 3).
InterPro; IPR001442; Procollagn4_C.	1604	CONFLICT	1604	E -> D (IN REF. 3).
Pfam; PF01413; C4; 2.	1682	CONFLICT	1682	P -> L (IN REF. 1 AND 3; AAA96216).
Pfam; PF01391; Collagen; 23.	1758	CONFLICT	1758	AA; 167750 MW; 97EE3F3D9B2D2AC5 CRC64;
ProDom; PD000007; Clg_helix; 6.	229	VARSPLIC	229	264
ProDom; PD003923; ProcollagnC4; 1.	248	VARSPLIC	248	264
SMART; SM00111; C4; 2.	258	VARSPLIC	258	264
Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;				
Alternative splicing; Glycoprotein; Signal.				
POTENTIAL.	1		26	
SIGNAL	27		1758	COLLAGEN ALPHA 2(IV) CHAIN.
CHAIN	27		42	7S DOMAIN.
DOMAIN	42		1527	TRIPLE-HELICAL REGION.
DOMAIN	1528		1758	NONHELICAL REGION (NCL).
DISULFID	1546		1635	OR 1632 (BY SIMILARITY).
DISULFID	1579		1632	OR 1635 (BY SIMILARITY).
DISULFID	1591		1597	BY SIMILARITY.
DISULFID	1654		1750	OR 1747 (BY SIMILARITY).
DISULFID	1688		1747	OR 1750 (BY SIMILARITY).
DISULFID	1700		1707	BY SIMILARITY.
CARBOHYD	248		248	O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
VARSPPLIC	229		264	GDGSGVPPGPGPIASTGSGIVSGVPRGNPGKGDK -> G DIGAMGAPGPPPIASTGSGITGKGLGKGEK (in isoform II).
FT	48		48	FT/ID: 001160.
FT	366		366	G -> E (IN MN114; 73% LETHAL).
FT	570		570	A -> T (IN MN126; 100% LETHAL).
FT	588		588	G -> E (IN MN109; 37% LETHAL).
FT	597		597	G -> R (IN MN103 AND MN151; 96% LETHAL).
FT	690		690	G -> R (IN MN132; 50% LETHAL).
FT	690		690	G -> R (IN MN101; 100% LETHAL).
FT	737		737	G -> E (IN MN129; 100% LETHAL).
FT	737		737	G -> E (IN MN143; 100% LETHAL).
FT	877		877	G -> R (IN G30; 90% LETHAL).
FT	904		904	G -> R (IN E1470; 94% LETHAL).
FT	1003		1003	G -> E (IN MN139; 20% LETHAL).
FT	1125		1125	G -> D (IN G25; 2% LETHAL).
FT	1152		1152	G -> D (IN G17 AND B246; 9% LETHAL).
FT	1286		1286	G -> D (IN REF. 3).
FT	1604		1604	E -> D (IN REF. 3).
FT	1682		1682	P -> L (IN REF. 1 AND 3; AAA96216).
FT	1758		1758	AA; 167750 MW; 97EE3F3D9B2D2AC5 CRC64;
FT	229		229	264
FT	248		248	264
FT	258		258	264
FT	264		264	264

```

P27393;
01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
Collagen alpha 2(IV) chain precursor.
Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
Ascariidae; Ascaris.
NCBI_TaxID=6253;
[1]
SEQUENCE FROM N.A. (ISOFORMS I AND II).
MEDLINE=91340768; PubMed=1714307;
Petitt J., Kingston I.B.;
and the complete primary structure of a nematode alpha 2(IV) collagen
and the partial structural organization of its gene.";
J. Biol. Chem. 266:16143-16156(1991).
-I- FUNCTION: Collagen type IV is specific for basement membranes.
-I- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
Type IV collagen forms a mesh-like network linked through
intermolecular interactions between 7S domains and between NC1
domains.
--I- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=i;
IsoId=P27393-1; Sequence=Displayed;
Name=ii;
IsoId=P27393-2; Sequence=VSP_001159;
DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
-I- PMW: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-I- PMW: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@sb-sib.ch).
-----
EMBL; M67507; AAA18014.1; -.
PIR; S16366; S16366.
InterPro; IPRO08161; Clg_helix.
InterPro; IPRO08160; Collagen.
InterPro; IPRO01442; Procollagn4_C.
Pfam; PF01413; C4; 2.
Pfam; PF01391; Collagen; 25.
ProDom; PD000007; Clg_helix; 6.
ProDom; PD003923; ProcollagnC4; 1.
SMART; SM00111; C4; 2.
Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
Alternative splicing; Glycoprotein; Signal.
FT FT 26 POTENTIAL.
CHAIN 1 27 COLLAGEN ALPHA 2(IV) CHAIN.
DOMAIN 27 42 7S DOMAIN.
DOMAIN 43 1529 TRIPLE-HELICAL REGION.
DOMAIN 1530 1763 NONHELIICAL REGION (NC1).
DISULFID 1548 1637 OR 1634 (BY SIMILARITY).
DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
DISULFID 1593 1599 BY SIMILARITY.
DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
DISULFID 1656 1752 OR 1752 (BY SIMILARITY).
DISULFID 1690 1749 BY SIMILARITY.
CARBOHYD 126 126 N-LINKED (GLCNAC.. ) (POTENTIAL).
CARBOHYD 249 249 O-LINKED (XYL.. ) (GLYCOSAMINOGLYCAN
(IN ISOFORM II) (POTENTIAL)).

```





RT "Papaya (Carica papaya) lysozyme is a member of the family 19 (basic, class II) chitinases.";  
RL J. Mol. Evol. 49:819-821 (1999).  
RN [2]  
RP SEQUENCE OF 1-5 AND 23-25.  
RC TISSUE=Leaf;  
RX MEDLINE=69130229; PubMed=5773045;  
RA Howard J.B., Glazer A.N.;  
RT "Papaya lysozyme. Terminal sequences and enzymatic properties.";  
RL J. Biol. Chem. 244:1399-1409 (1969).  
CC -!- FUNCTION: Bifunctional enzyme with lysozyme/chitinase activity.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-acetyl-D-glucosamine polymers of chitin.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan heteropolymers of the prokaryotes cell walls.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- SIMILARITY: Belongs to family 19 of glycosyl hydrolases.  
DR InterPro: IPR000726; Glyco\_hydro\_19  
DR PROSITE: PS00773; CHITINASE\_19\_1; PARTIAL.  
DR PROSITE: PS00774; CHITINASE\_19\_2; PARTIAL.  
KW Hydrolase; Glycosidase; Chitin degradation; Multifunctional enzyme.  
FT CONFLICT 3 3 E -> S (IN REF. 2).  
FT NON\_CONS 22 23  
FT NON\_TER 25 25  
SQ SEQUENCE 25 AA; 2877 MW; 33BA3F018F33ACD6 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 25;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 230 EXIISR 235  
DB 3 EXIISR 8  
  
RESULT 81  
PSAI CHAGL  
ID PSAL CHAGL STANDARD; PRT; 36 AA.  
AC Q8MX55;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Photosystem I reaction center subunit VIII (PSI-I).  
GN PSAL  
OS Chaetosphaeeridium globosum.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Coleochaetales;  
OC Chaetosphaeeridiaceae; Chaetosphaeeridium.  
OX NCBI\_TaxID=96477;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M311;  
RX MEDLINE=22177139; PubMed=12161550;  
RA Turmel M., Otis C., Lemeux C.;  
RT "The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeeridium globosum: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:11275-11280 (2002).  
RL  
CC -!- FUNCTION: May help in the organization of the psal subunit (By similarity).  
CC -!- SIMILARITY: Belongs to the psal family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; AF494278; AAM96535.1; -.  
DR HAMAP; MF\_00431; -; 1.  
DR InterPro; IPR001302; PSI\_8.  
DR Pfam; PF00796; PSI\_8; 1.  
KW Photosystem I; Photosynthesis; Transmembrane; Chloroplast.  
FT TRANSMEM 5  
FT POTENTIAL 27  
SQ SEQUENCE 36 AA; 3876 MW; 963F419C07BD405D CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 36;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 128 PAIAIA 133  
DB 19 PAIAIA 24  
  
RESULT 82  
LA89 LACAC  
ID LA89 LACAC STANDARD; PRT; 46 AA.  
AC Q48501; Q9RSH2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Bacteriocin acidocin 8912 precursor.  
GN ACDT.  
OS Lactobacillus acidophilus.  
OG Plasmid PLA103.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1579;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK8912;  
RX MEDLINE=96140009; PubMed=8554765;  
RA Kanatani K., Tahara T., Oshimura M., Sano K., Umezawa C.;  
RT "Cloning and nucleotide sequence of the gene for acidocin 8912, a bacteriocin from Lactobacillus acidophilus TK8912.";  
RT Lett. Appl. Microbiol. 21:384-386 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK8912;  
RX Nakachi I.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE OF 21-44.  
RC STRAIN=TK8912;  
RX MEDLINE=93005074; PubMed=1368836;  
RA Tahara T., Kanatani K., Yoshida K., Miura H., Sakamoto M., Oshimura M.;  
RT "Purification and some properties of acidocin 8912, a novel bacteriocin produced by Lactobacillus acidophilus TK8912.";  
RT Biosci. Biotechnol. Biochem. 56:1212-1215 (1992).  
RL  
CC -!- FUNCTION: Has a bactericidal effect on sensitive cells but not a bacteriolytic effect.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; D43626; BAA07737.1; -.  
DR EMBL; AB081463; BAB86322.1; -.  
KW Antibiotic; Bacteriocin; Plasmid.  
FT PROPEP 1 20  
FT CHAIN 21 46 BACTERIOCIN ACIDOCIN 8912.  
FT CONFLICT 33 33 W -> R (IN REF. 3).  
FT CONFLICT 46 AA; 5331 MW; EAF910D04D2AC3E8 CRC64;  
SQ SEQUENCE

```
Query Match          2.5%; Score 6; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SLWKG 155
Db 31 SLWKG 36

RESULT 83
PSBZ_CYAPA
ID PSBZ_CYAPA STANDARD; PRT; 65 AA.
AC P17159;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Photosystem II reaction center Z protein.
GN PSBZ OR YCF9.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Rukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
[1]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RX MEDLINE=91346714; PubMed=2129400;
RA Erward J.L., Kuntz M., Weil J.H.;
RT "An ORF potentially encoding a 6.5 kDa hydrophobic protein in
RT chloroplasts is also present in the cyanellar genome of Cyanophora
RT paradoxa.";
RL Plant Mol. Biol. 15:779-781(1990).
[2]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
[3]
SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.B.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- FUNCTION: Controls the interaction of photosystem II (PSII) cores
CC with the light-harvesting antenna (By similarity).
CC -1- SUBCELLULAR LOCATION: Cellular thylakoid membrane; associated with
CC the photosystem II complex (By similarity).
CC -1- SIMILARITY: Belongs to the psbZ family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51421; CAA35786.1; -
CC EMBL; U30821; AAA81235.1; -
CC PIR; S14712; S14712.
CC HAMAP; MF_00644; -; 1
CC InterPro; IPR002644; Ycf9_struc.
CC Pfam; PF01737; YCF9; 1.
CC Photosynthesis; Photosystem II; Reaction center; Thylakoid;
CC Transmembrane; Cyanelle.
CC TRANSMEM 8 28
CC POTENTIAL.

FT TRANSMEM 41 61 POTENTIAL.
SQ SEQUENCE 65 AA; 6853 MW; 9FE5CD75B895623C CRC64;

Query Match          2.5%; Score 6; DB 1; Length 65;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPG 176
Db 26 ALASPG 31

RESULT 84
CSRA_XYLFA
ID CSRA_XYLFA STANDARD; PRT; 71 AA.
AC Q9PH21; Q97F42;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbon storage regulator homolog.
GN CSRA OR XF0125 OR PD0095.
OS Xylella fastidiosa, and
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371, 183190;
[1]
SEQUENCE FROM N.A.
RP STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Falcinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite J.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
[2]
SEQUENCE FROM N.A.
RC STRAIN=Temeculal / ATCC 700964;
RX MEDLINE=2241331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Moon D.H.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., de Silva F.R., Tsai S.M.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.S., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
```

RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,  
RA Kitajima J.P.;  
RT "Comparative analyses of the complete genome sequences of Pierce's  
RT disease and citrus variegated chlorosis strains of Xylella  
RT fastidiosa";  
RL J. Bacteriol. 185:1018-1026(2003).  
CC -!- FUNCTION: Could accelerate the degradation of some genes  
CC transcripts potentially through selective RNA binding (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the csrA family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; AE003866; AAF82938.1; ALT\_INIT.  
DR EMBL; AE012553; AAO27995.1; -.  
DR FIC; D82844; D82844.  
DR HAMAP; MF\_00167; -; 1.  
DR InterPro; IPR003751; CsrA.  
DR Pfam; PF02599; CsrA; 1.  
DR ProDom; PD009007; CsrA; 1.  
DR TIGRFAMs; TIGR00202; csrA; 1.  
KW RNA-binding; Complete proteome.  
FT DOMAIN 11 45  
FT SEQUENCE 71 AA; 7672 MW; FBA182A40CE72AD CRC64;  
SQ

Query Match 2.5%; Score 6; DB 1; Length 71;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGS 11  
DB 64 RGDGS 69

RESULT 85  
C555\_METCA STANDARD; PRT; - 96 AA.  
AC P04369;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Cytochrome c-555 (C555).  
OS Methylococcus capsulatus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Methylococcales;  
OC Methylococcaceae; Methylococcus.  
OX NCBI\_TaxID=414;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=Bath / NCIB 11132;  
RX MEDLINE=86158741; PubMed=3006666;  
RA Ambler R.P., Dalton H., Meyer T.E., Bartsch R.G., Kamen M.D.;  
RT "The amino acid sequence of cytochrome c-555 from the  
RT methane-oxidizing bacterium Methylococcus capsulatus";  
RL Biochem. J. 233:333-337(1986).  
DR FIC; A23321; CCMPS5.  
DR InterPro; IPR003088; Cyt\_C1.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR Pfam; PF00034; Cytochrome C; 1.  
DR PROSITE; PS00190; CYTOCHROME C; 1.  
KW Electron transport; Photosynthesis; Heme.  
FT BINDING 19 19 HEME (COVALENT).  
FT BINDING 22 22 HEME (COVALENT).  
FT METAL 23 23 IRON (HEME AXIAL LIGAND).  
FT METAL 59 59 IRON (HEME AXIAL LIGAND).  
SQ SEQUENCE 96 AA; 10506 MW; 17B5DCF79535A585 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 96;

Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AGELEK 231  
DB 89 AGELEK 94

RESULT 86  
RSN\_MOUSE STANDARD; PRT; 114 AA.  
ID RSN\_MOUSE  
AC Q99P87;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Resistin precursor (Cysteine-rich secreted protein FIZZ3) (Adipose  
DE tissue-specific secretory factor) (ADSF) (Adipose-specific cysteine-  
DE rich secreted protein A12-alpha).  
GN RETN OR FIZZ3  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RX MEDLINE=21069045; PubMed=11201732;  
RA Steppan C.M., Bailey S.T., Bhat S., Brown E.J., Banerjee R.R.,  
RA Wright C.M., Patel H.R., Ahima R.S., Lazar M.A.;  
RT "The hormone resistin links obesity to diabetes";  
RL Nature 409:307-312(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rajala M.W., Scherer P.E.;  
RT "Identification of a novel adipose-specific cysteine-rich secreted  
RT protein";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SUBUNIT.  
RX MEDLINE=21336653; PubMed=11358969;  
RA Banerjee R.R., Lazar M.A.;  
RT "Dimerization of resistin and resistin-like molecules is determined by  
RT a single cysteine";  
RL J. Biol. Chem. 276:25970-25973(2001).  
CC -!- FUNCTION: Hormone that seems to suppress insulin ability to  
CC stimulate glucose uptake into adipose cells. Potentially  
CC links obesity to diabetes.  
CC -!- SUBUNIT: Homodimer; disulfide-linked.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed in white but not brown adipose

tissue in a variety of organs.  
-!- PTM: 5 disulfide bonds are present (probable).  
-!- SIMILARITY: Belongs to the resistin/Fizz family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF323080; RAGS9823.1; -;  
DR EMBL; AF290870; AAK3102.1; -;  
DR EMBL; BC051196; AAH51196.1; -;  
DR MGD; MGI:1888506; Rtn.  
DR GO; GO:0005576; C:extracellular; IDA.  
DR Hormone; Signal; Diabetes mellitus; Obesity.  
FT SIGNAL 1 20  
FT CHAIN 21 114 RESISTIN.  
FT DISULFID 26 26 INTERCHAIN (PROBABLE).  
SQ SEQUENCE 114 AA; 12491 MW; D44930E51D3F22C8 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 114;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 33 SCPEPT 38  
Db 66 SOPEPT 71  
|||||  
-----  
RESULT 87  
ID CN4D MOUSE STANDARD; PRT; 116 AA.  
AC Q01063;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE CAMP-specific 3',5'-cyclic phosphodiesterase 4D (EC 3.1.4.17) (DPDE3) (fragment).  
DE PDB4D.  
GN PDB4D.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92406782; PubMed=1326532;  
RA Repaske D.R., Swinnen J.V., Jin S.L.C., van Wyk J.J., Conti M.;  
RT "A polymerase chain reaction strategy to identify and clone cyclic nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA encoding the 63-kDa calmodulin-dependent phosphodiesterase.";  
RL J. Biol. Chem. 267:18683-18688 (1992).  
CC -!- CATALYTIC ACTIVITY: Adenosine 3',5'-cyclic phosphate + H(2)O = adenosine 5'-phosphate.  
CC -!- ENZYME REGULATION: Inhibited by rolipram.  
CC -!- PATHWAY: Cyclic nucleotide metabolism.  
CC -!- SIMILARITY: Belongs to the cyclic nucleotide phosphodiesterase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M94541; AAA37368.1; -;  
DR MGD; MGI:99555; Pde4d.  
DR InterPro; IPR002073; PDEase.  
DR Pfam; PF00233; PDEase; 1.  
RP SEQUENCE FROM N.A.

PRINTS; PR00387; PD1ESTERASE1.  
DR PROSITE; PS00126; PDEASE\_1; 1.  
KW Hydrolase; CAMP; Alternative splicing; Multigene family.  
FT NON\_TER 1  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 12928 MW; D217AF0F432611B2 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 97 LSTPAL 102  
Db 18 LSTPAL 23  
|||||  
-----  
RESULT 88  
ID Y128 SYNPE STANDARD; PRT; 119 AA.  
AC P05677;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 12.8 kDa protein in 16S rRNA gene region.  
OS Synecococcus sp. (strain POC 6301) (Anacystis nidulans).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
OX NCBI\_TaxID=1139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kumano M., Tomioka N., Shinozaki K., Sugitara M.;  
RT "Analysis of the promoter region in the rRNA operon from a blue-green alga, Anacystis nidulans 6301.";  
RL Mol. Gen. Genet. 202:173-178 (1986).  
CC -----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X03538; CAA27241.1; -;  
DR PIR; S10914; S10914.  
DR InterPro; IPR004843; M-pdestrase.  
DR Pfam; PF00149; Metallophos; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 119 AA; 12867 MW; F1032C437D5DFEF6 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 119;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 128 PAIAIA 133  
Db 73 PAIAIA 78  
|||||  
-----  
RESULT 89  
ID YWJ0 YEAST STANDARD; PRT; 128 AA.  
AC Q04501;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Very hypothetical 15.0 kDa protein in RPM2-TUB1 intergenic region. YML090W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.

```

RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Iye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93 (1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z46660; CAA86648.1; -.
CC PIR; S49537; S49637.
CC GenOnline; 142621; -.
CC SGD; S0004555; YML090W.
CC
CC Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT SEQUENCE 128 AA; 15034 MW; A4AE42CDFA441B6B CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 45 PSFLFV 50
Db 4 PSFLFV 9
RESULT 90
ID CAL2_MOUSE STANDARD; PRT; 130 AA.
AC Q99MP3;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calcitonin gene-related peptide II precursor (CGRP-II) (Beta-type
DE CGRP).
DE CALCB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21604266; PubMed=11761712;
RA Thomas P.M., Nasonkin I., Zhang H., Gagel R.F., Cote G.J.;
RT "Structure of the mouse calcitonin/calcitonin gene-related peptide
RT alpha and beta genes."
RL DNA Seq. 12:131-135 (2001).
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the calcitonin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF325526; AAK16431.1; -.
CC EMBL; AF325524; AAK16431.1; JOINED.
DR MGD; MGI:2151254; Calcb.
DR InterPro; IPR001693; Calcitonin-like.
DR InterPro; IPR002163; Calcitonin B.
DR Pfam; PF00214; Calc. CGRP IAPP; 1.
DR PRINTS; PR00817; CALCITONINB.
DR SMART; SM00113; CALCITONIN; 1.
DR PROSITE; PS00258; CALCITONIN; 1.
KW Cleavage on pair of basic residues; Amidation; Hormone; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 82 BY SIMILARITY.
FT PEPTIDE 84 120 CALCITONIN GENE-RELATED PEPTIDE II.
FT PROPEP 127 130 BY SIMILARITY.
FT DISULFID 85 90 BY SIMILARITY.
FT MOD_RES 120 120 AMIDATION (G-121 PROVIDE AMIDE GROUP) (BY
FT SIMILARITY).
SQ SEQUENCE 130 AA; 14623 MW; 97299244E8F6C536 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 60 DLGTLG 65
Db 38 DLGTLG 43
RESULT 91
ID YAEJ_PSEPU STANDARD; PRT; 137 AA.
AC P45388;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical 15.2 kDa protein in pcau 3 region.
DE Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRS2000;
RX MEDLINE=92325057; PubMed=1624453;
RA Parales R.E., Harwood C.S.;
RT "Characterization of the genes encoding beta-ketoadipate: succinyl-
RT coenzyme A transferase in Pseudomonas putida."
RL J. Bacteriol. 174:4657-4666 (1992).
CC -1- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
CC factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88763; -; NOT ANNOTATED CDS.
DR InterPro; IPR000352; Pep_rel_factor_I.
DR Pfam; PF00472; RF-1; 1.
DR PROSITE; PS00745; RF_PROK_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 137 AA; 15181 MW; BDDAF2B68986A2EC CRC64;
Query Match 2.5%; Score 6; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 222 STVAG 227

```

CC - I- FUNCTION: Troponin is the central regulatory protein of striated muscle contraction.





```
RT of stone formation.";  
RL EMBO J. 15:2678-2684(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (1.30 ANGSTROMS) OF 23-166.  
RX MEDLINE=20092874; PubMed=10625646;  
RA Gerbaud V., Pignol D., Loret E., Bertrand J.A., Berland Y.,  
RA Fontcilla-Camps J.C., Canselier J.P., Gabas N., Verdier J.M.;  
RT Mechanism of calcite crystal growth inhibition by the N-terminal  
RT undecapeptide of lithostathine.";  
RL J. Biol. Chem. 275:1057-1064(2000).  
RN [16]  
RN STRUCTURE BY NMR OF 34-164.  
RX MEDLINE=97120677; PubMed=8961348;  
RA Patard L., Stoven V., Gharib B., Bontems F., Lallemand J.-Y.,  
RA de Reggi M.;  
RT What function for human lithostathine?: structural investigations by  
RT three-dimensional structure modeling and high-resolution NMR  
RT spectroscopy.";  
RL Protein Eng. 9:949-957(1996).  
CC -!- FUNCTION: Might act as an inhibitor of spontaneous calcium  
CC carbonate precipitation. May be associated with neuronal  
CC sprouting in brain, and with brain and pancreas regeneration.  
CC -!- TISSUE SPECIFICITY: In pancreatic acinar cells and, in lower  
CC levels, in brain.  
CC -!- DEVELOPMENTAL STAGE: High expression levels in fetal and infant  
CC brains; much lower in adult brains.  
CC -!- DISEASE: Alzheimer's disease and Down's syndrome patients show  
CC enhanced expression of RSP-related transcripts and intraneuronal  
CC accumulation of RSP-like proteins in their brains.  
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M27130; AAA60546.1; -  
DR EMBL; M27189; AAA60545.1; -  
DR EMBL; M18963; AAA36558.1; -  
DR EMBL; J05412; AAA36559.1; -  
DR EMBL; AF172331; AAD51330.1; -  
DR EMBL; BC005350; AAA05350.1; -  
DR PIR; A35197; RGHU1A.  
DR PIR; A45751; A45751.  
DR PDB; 1LIT; 11-JAN-97.  
DR PDB; 1QDD; 24-JAN-01.  
DR Genew; HGNC:19951; REG1A.  
DR MIM; 167770; -  
DR MIM; 167800; -  
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.  
DR InterPro; IPR002353; Antifreeze1.  
DR InterPro; IPR001304; Lectin.C  
DR InterPro; IPR003990; Pancreat1s_ac.  
DR Pfam; PF00059; lectin_C; 1.  
DR PRINTS; PR00356; ANTIFREEZE1.  
DR PRINTS; PR01504; PNCRAT1SAP.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C TYPE LECTIN 1; 1.  
DR PROSITE; PS50041; C TYPE LECTIN 2; 1.  
KW Glycoprotein; Signal; Alzheimer's disease; Down's syndrome; Lectin;  
KW 3D-structure; Pyrrolidone carboxylic acid.  
FT SIGNAL 1 22  
Query Match 2.5%; Score 6; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 33 SCEPT 38  
D6 35 SCEPT 40
```

```
RESULT 96  
LITE_HUMAN  
ID LITE_HUMAN STANDARD; PRT; 166 AA.  
AC P48304;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lithostathine 1 beta precursor (Regenerating protein I beta).  
GN REG1B OR REG1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=94153997; PubMed=8110835;  
RA Morizumi S., Watanabe T., Unno M., Nakagawara K.I., Suzuki Y.,  
RA Miyashita H., Yonekura H., Okamoto H.;  
RT Isolation, structural determination and expression of a novel reg  
RT gene, human regi beta.";  
RL Biochim. Biophys. Acta 1217:199-202(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=93351647; PubMed=8348956;  
RA Bartoli C., Gharib B., Giorgi D., Sansonetti A., Dagorn J.-C.,  
RA Berge-Lefranc U.;  
RT "A gene homologous to the reg gene is expressed in the human  
RT pancreas.";  
RL FEBS Lett. 327:289-293(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Abramson R.D., Mullany S.J.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Mckernan K.J., Malek J.A., Gunaratne P.H.,  
RA Bosak S.A., McWray K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
RN [4]  
RP CARBOHYDRATE-LINKAGE SITE.  
RX MEDLINE=95331286; PubMed=7607222;  
RA De Reggi M., Capon C., Gharib B., Wieruszski J.M., Michel R.,  
RA Fournet B.;  
RT The glycan moiety of human pancreatic lithostathine. Structure  
RT characterization and possible pathophysiological implications.";  
RL Eur. J. Biochem. 230:503-510(1995).  
CC -!- FUNCTION: Might act as an inhibitor of spontaneous calcium  
CC carbonate precipitation. May be associated with neuronal sprouting  
CC in brain, and with brain and pancreas regeneration.  
CC -!- PTM: ALL O-LINKED GLYCANS CONSIST OF GAL-GLCNAc-GAL-GALNAc  
CC TETRASACCHARIDE CORE AND GET ELONGATED (MICROHETEROGENEITY).  
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D17291; BAA04124.1; -  
DR EMBL; D16816; BAA04091.1; -  
DR EMBL; L08010; AAA18204.1; -  
DR EMBL; BC027895; AAH27895.1; -  
DR PIR; S34591; RGHU1B.  
DR HSSP; P05451; LLIT.  
DR Genew; HGNC:9952; REG1B.  
DR MIM; 167771; -  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR InterPro; IPR001304; Lectin\_C.  
DR InterPro; IPR003990; Pancreatins\_ac.  
DR Pfam; PF00059; lectin\_c; 1.  
DR PRINTS; PR01504; PNCREATITSP.  
DR SMART; SM00034; CLECT.1.  
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.  
DR GlycoProtein; Signal; Lectin; Pyroglutamate carboxylic acid.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 166 LITHOSTATHINE 1 BETA.  
FT DOMAIN 34 164 C-TYPE LECTIN (LONG FORM).  
FT MOD\_RES 23 23 PYROGLUTAMATE CARBOXYLIC ACID  
FT (BY SIMILARITY).  
FT CARBOHYD 27 27 O-LINKED (GALNAc... ) (MUCIN TYPE).  
FT DISULFID 36 47 BY SIMILARITY.  
FT DISULFID 64 162 BY SIMILARITY.  
FT DISULFID 137 154 BY SIMILARITY.  
SQ SEQUENCE 166 AA; 18665 MW; D1DC20E11AE5DDEB CRC64;

Query Match 2.5%; Score 6; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SCSPT 38  
DB 35 SCSPT 40  
|||||

RESULT 97  
VB03 VACCV STANDARD; PRT; 167 AA.  
AC Q0126;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Protein B3.  
GN B3.  
OS Vaccinia virus (strain WR).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
OX NCBI\_TaxID=10254;  
FN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91259063; PubMed=2045793;  
RA Smith G.L., Chan Y.S., Howard S.T.;  
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near  
RT the right inverted terminal repeat";  
RL J. Gen. Virol. 72:1349-1376(1991).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D11079; BAA01833.1; -

DR PIR; JQ1797; JQ1797.  
SQ SEQUENCE 167 AA; 19410 MW; 82AF46891A7768D7 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 167;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 220 IPSTVK 225  
DB 39 IPSTVK 44  
|||||

RESULT 98  
RECU UREPA STANDARD; PRT; 170 AA.  
ID RECU UREPA  
AC Q9PQJ4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Recombination protein U homolog.  
GN RECU OR U297.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
FN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Serovar 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
RT urealyticum";  
RL Nature 407:757-762(2000).  
CC -!- FUNCTION: Required for DNA repair and intramolecular  
CC recombination. Seems also to be required for chromosome  
CC segregation (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the recU family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE002127; AAF30706.1; -  
DR HAMAP; MF 00130; -; 1.  
DR InterPro; IPR004612; RecU.  
DR Pfam; PF03838; RecU; 1.  
DR DNA repair; DNA recombination; Complete proteome.  
SQ SEQUENCE 170 AA; 20527 MW; 892315A8C36933DD CRC64;

Query Match 2.5%; Score 6; DB 1; Length 170;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SGFSFL 48  
DB 99 SGFSFL 104  
|||||

RESULT 99  
RL10 THEAC STANDARD; PRT; 176 AA.  
ID RL10 THEAC  
AC Q9HJB3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L10e.  
GN RPL10E OR TAI057.  
OS Thermoplasma acidophilum.

```
OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatatales;
OC Thermoplasmatataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Grail W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Newes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
CC -1- SIMILARITY: Belongs to the L10e family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR HAMAP; MF 00448; -; 1.
DR InterPro; IPR001197; Ribosomal L10E.
DR Pfam; PF00826; Ribosomal L10e; 1.
DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 176 AA; 19377 MW; E284F9F14A187B46 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 FVQGNQ 54
DB 34 FVQGNQ 39

RESULT 101
INAA BOVIN
ID INAA BOVIN STANDARD; PRT; 189 AA.
AC P05007;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Interferon alpha-A precursor.
GN IFNAA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85182698; PubMed=3886658;
RA Velan B., Cohen S., Grosfeld H., Leitner M., Shafferman A.;
RT "Bovine interferon alpha genes. Structure and expression."
RL J. Biol. Chem. 260:5458-5504(1985)
CC -1- FUNCTION: PRODUCED BY MACROPHAGES, IFN-ALPHA HAVE ANTIVIRAL
CC ACTIVITIES. INTERFERON STIMULATES THE PRODUCTION OF TWO ENZYMES:
CC A PROTEIN KINASE AND AN OLIGODENYLATE SYNTHETASE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the alpha/beta interferon family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Db          95 GSPATW 100

RESULT 102
SODN STAAM
ID SODN STAAM STANDARD; PRT; 199 AA.
AC Q99X82;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Superoxide dismutase [Mn/Fe] 2 (EC 1.15.1.1).
GN SOD2 OR SAV0133 OR SA0128 OR MW0107.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcaceae.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kunara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RA "whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=WM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Aeono K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 manganese or iron ion per subunit (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003358; BAB56295.1; -
CC EMBL; AP003129; BAB41348.1; -
CC EMBL; AP004822; BAB93972.1; -
CC PIR; A89774; A89774.
CC HSSP; P09214; 1MNG.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PRODOM; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.
KW METAL 27 27 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 81 81 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 161 161 MANGANESE OR IRON (BY SIMILARITY).
FT METAL 161 161 MANGANESE OR IRON (BY SIMILARITY).

FT METAL 165 165 MANGANESE OR IRON (BY SIMILARITY).
SQ SEQUENCE 199 AA; 23041 MW; 388566FB9943C635 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYI 119
Db 14 ALEPYI 19

RESULT 103
SODM STRPN
ID SODM STRPN STANDARD; PRT; 200 AA.
AC Q59949; Q33757; Q54268; Q54269; Q9R3B8; Q9R3B8; Q9S176;
AC Q9S177; Q9S447;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] (EC 1.15.1.1).
GN SODA OR SP0766 OR SPO674.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=D39 / NCTC 7466 / Serotype 2;
RX MEDLINE=20231823; PubMed=10768978;
RA Yesilkaya H., Kadioglu A., Gingles N., Alexander J.E., Mitchell T.J.,
RA Andrew P.W.;
RT "Role of manganese-containing superoxide dismutase in oxidative stress
RT and virulence of Streptococcus pneumoniae.";
RL Infect. Immun. 68:2819-2826(2000).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Whitt O., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White C., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapf E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21423245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Bargett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Legace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nickas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RN [4]
SEQUENCE OF 17-161 FROM N.A.
RP STRAIN=NEW667;
RX MEDLINE=96045282; PubMed=7557308;
RA Poyart C., Berche P., Trieu-Cuot P.;
RT "Characterization of superoxide dismutase genes from Gram-positive
RT bacteria by polymerase chain reaction using degenerate primers.";
```

RL FEMS Microbiol. Lett. 131:41-45(1995).

RN [5]

RP SEQUENCE OF 17-161 FROM N.A.

RC STRAIN=Various strains;

RX MEDLINE=98092214; PubMed=9431917;

RY Poyart C., Quesne G., Coulon S., Berche P., Trieu-Cuot P.;

RT "Identification of streptococci to species level by sequencing the

RT gene encoding the manganese-dependent superoxide dismutase.";

RL J. Clin. Microbiol. 36:41-47(1998).

RL [6]

RP SEQUENCE OF 32-153 FROM N.A.

RC STRAIN=653, 661, 872, 1293, 1454, 1510, 1565, 1639, 3051, 3203,

RC GTC361T / NCTC 7465T, YK-5, YK-11, YK-12, YK-14, and YK-20;

RX MEDLINE=99445202; PubMed=10517614;

RY Kawamura Y., Whitley R.A., Shu S.E., Ezaki T., Hardie J.M.;

RT "Genetic approaches to the identification of the mitis group within

RT the genus Streptococcus.";

RL Microbiology 145:260S-2613(1999).

CC -!- FUNCTION: Destroys radicals which are normally produced within the

CC cells and which are toxic to biological systems. May play a

CC critical role against oxidative stress, affecting both the survival

CC and the virulence of S.pneumoniae.

CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).

CC -!- SUBUNIT: Homodimer (By similarity).

CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR ENBL; AF162564; AA050778.1; -

DR ENBL; AE007384; AAK74904.1; -

DR ENBL; AE008445; AAK99478.1; -

DR ENBL; Z49246; CAA89213.1; -

DR ENBL; Z95914; CAA09367.1; -

DR ENBL; Z99200; CABI6344.1; -

DR ENBL; Z99201; CABI6345.1; -

DR ENBL; Z99202; CABI6346.1; -

DR ENBL; Z99203; CABI6347.1; -

DR ENBL; Z99204; CABI6348.1; -

DR ENBL; Z99205; CABI6349.1; -

DR ENBL; Z99206; CABI6350.1; -

DR ENBL; AB021544; BAA85492.1; -

DR ENBL; AB021605; BAA85553.1; -

DR ENBL; AB021606; BAA85554.1; -

DR ENBL; AB021607; BAA85555.1; -

DR ENBL; AB021608; BAA85556.1; -

DR ENBL; AB021609; BAA85557.1; -

DR ENBL; AB021610; BAA85558.1; -

DR ENBL; AB021611; BAA85559.1; -

DR ENBL; AB021612; BAA85560.1; -

DR ENBL; AB021613; BAA85561.1; -

DR ENBL; AB021614; BAA85562.1; -

DR ENBL; AB021615; BAA85563.1; -

DR ENBL; AB021616; BAA85564.1; -

DR ENBL; AB021617; BAA85565.1; -

DR ENBL; AB021618; BAA85566.1; -

DR ENBL; AB021619; BAA85567.1; -

DR PR; S54795; S54795.

DR HSP; P09214; INMG.

DR TGR; SP0766; -

DR InterPro; IPR001189; SODismutase.

DR Pfam; PF00081; sdcfe; 1.

DR Pfam; PF02777; sdcfe; 1.

DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom; PD000475; SODismutase; 1.

DR PROSITE; PS00086; SOD\_MN; 1.

KW Oxidoreductase; Metal-binding; Manganese; Complete proteome.

FT INIT\_MET 0 BY SIMILARITY.

FT METAL 26 MANGANESE (BY SIMILARITY).

FT METAL 80 MANGANESE (BY SIMILARITY).

FT METAL 162 MANGANESE (BY SIMILARITY).

FT METAL 166 MANGANESE (BY SIMILARITY).

FT VARIANT 41 A -> V (IN STRAIN NEM1278).

FT VARIANT 60 ESIPA -> DLSQH (IN STRAINS NEM667 AND

FT D39).

FT VARIANT 99 A -> T (IN STRAIN 1510).

FT VARIANT 109 F -> L (IN STRAIN NEM1122).

FT VARIANT 142 T -> I (IN STRAIN 1510).

FT VARIANT 147 Q -> P (IN STRAINS 1293, 1454, 1565,

FT 1639, 3051 AND 3203).

FT VARIANT 147 Q -> L (IN STRAIN 872).

FT VARIANT 151 I -> F (IN STRAIN 661).

FT VARIANT 151 I -> F (IN STRAIN 661).

SQ SEQUENCE 200 AA; 22266 MW; DB2F4E666830658F CRC64;

Query Match 2.5%; Score 6; DB 1; Length 200;

Best Local Similarity 100.0%; Pred. No. 1.7e-02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALPEYI 119

Db 13 ALPEYI 18

RESULT 104

CTD6 HUMAN

ID\_CTD6 HUMAN STANDARD; PRT; 202 AA.

AC Q9BFL; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C20orf136.

GN C20ORF136

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

FN [1]

RP SEQUENCE FROM N.A.

EX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagdley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Leharalain M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConnachie L.J., Melay K., McMurray A.A.,

RA Milne S.A., Mistry D., Moore W.J.F., Mullikin J.C., Nickerson I.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,

RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20.";

RL Nature 414:865-971(2001).

CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL118506; CAC28315.1; -.  
DR Genew; HGNC:16129; C200xf136.  
DR InterPro; IPR001660; SAM.  
DR Pfam; PF00536; SAM; 1.  
DR SMART; SM00454; SAM; 1.  
DR PROSITE; PSS0105; SAM\_DOMAIN; 1.  
KW Hypothetical protein.  
FT DOMAIN 118 184 SAM.  
SQ SEQUENCE 202 AA; 22770 MW; DF2678F090A3B946 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 202;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 ITGRAL 115  
|||||  
DB 147 ITGRAL 152

RESULT 105  
SODM\_STRMU STANDARD; PRT; 202 AA.  
AC P09738; Q59791;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).  
GN SODA OR SOD, OR SMO.629.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=92332426; PubMed=1321118;  
RA Nakayama K.;  
RT "Nucleotide sequence of Streptococcus mutans superoxide dismutase  
RT gene and isolation of insertion mutants.";  
RL J. Bacteriol. 174:4928-4934(1992).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
RX MEDLINE=2295063; PubMed=12397186;  
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
RT pathogen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

RN [3]  
RP SEQUENCE OF 1-22.  
RX MEDLINE=96250886; PubMed=3722201;  
RA Martin M.E., Byers B.R., Olson M.O.J., Salin M.L., Arceneaux J.E.L.,  
RA Tolbert C.;  
RT "A Streptococcus mutans superoxide dismutase that is active with  
RT either manganese or iron as a cofactor.";  
RL J. Biol. Chem. 261:9361-9367(1986).

CC -1- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems.

CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -1- COFACTOR: Binds 1 manganese or iron ion per subunit (By  
CC similarity).

CC -1- SIMILARITY: Belongs to the iron/manganese superoxide dismutase  
CC family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D01037; BAB86970.1; -.  
DR EMBL; AB014906; AAN58363.1; -.  
DR PIR; A42710; A42710.  
DR HSPSP; P00448; IVEW.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe\_1.  
DR Pfam; PF02777; sodfe\_C1.  
DR PRINTS; PR01703; MNSODISMUTASE.  
DR PROSITE; PS000475; SODismutase; 1.  
DR PROSITE; PS00088; SOD MN; 1.  
KW Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.

FT INIT MET 0  
FT METAL 26 26 MANGANESE OR IRON (BY SIMILARITY).  
FT METAL 80 80 MANGANESE OR IRON (BY SIMILARITY).  
FT METAL 162 162 MANGANESE OR IRON (BY SIMILARITY).  
FT METAL 166 166 MANGANESE OR IRON (BY SIMILARITY).  
FT CONFLICT 3 3 L -> T (IN REF. 3).  
SQ SEQUENCE 202 AA; 22494 MW; C0C853BE0032B541 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 202;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEFYI 119  
|||||  
DB 13 ALPEFYI 18

RESULT 106  
NU6M\_TRIRU STANDARD; PRT; 203 AA.  
ID NU6M\_TRIRU  
AC Q36836;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE NAH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).  
GN ND6 OR NADH6.  
OS Trichophyton rubrum.  
OC Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;  
OC Trichophyton.  
OX NCBI\_TaxID=5551;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IP 1817.89;  
RX MEDLINE=96132111; PubMed=8593686;  
RA de Bievre C., Dujon B.;  
RT "Organisation of the mitochondrial genome of Trichophyton rubrum. DNA  
RT sequence analysis of the ND4 gene, the ATPase subunit-6 gene, the  
RT ribosomal RNA small-subunit gene, the ND6 gene, the COXII gene, the  
RT ATPase subunit-8 gene and six tRNA genes that correspond respectively  
RT to the tyrosine, lysine, glutamine, asparagine, isoleucine and  
RT tryptophan isocodons.";  
RL Curr. Genet. 28:553-559(1995).

CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
DR EMBL; X88896; CAA61356.1; -.
DR PIR; S65033; S65033.
DR InterPro; IPR001457; Oxidored g3.
DR Pfam; PF00499; Oxidored g3; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 203 AA; 22656 MW; 046DAB3716E02802 CRC64;

Query Match          2.5%; Score 6; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 202 SYSFWL 207
DB 175 SYSFWL 180

RESULT 107
SOD2_PLEBO          STANDARD;          PRT; 206 AA.
ID _JAG_BACHD
AC P50059;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] 2 [EC 1.15.1.1].
GN SOD2.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxID=1184;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX 485;
RX MEDLINE=95164530; PubMed=7860607;
RA Campbell W.S., Laudenbach D.B.;
RT "Characterization of four superoxide dismutase genes from a filamentous cyanobacterium."
RL J. Bacteriol. 177:964-972(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- INDUCTION: BY METHYL VIOLOGEN, AND UNDER CONDITIONS OF IRON OR NITROGEN STRESS.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U17610; AAA69952.1; -.
CC HSP; P09214; LMNG.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe C; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC PRODOM; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Multigene family.
KW METAL.
FT METAL 27 27 MANGANESE (BY SIMILARITY).
FT METAL 82 82 MANGANESE (BY SIMILARITY).
FT METAL 165 165 MANGANESE (BY SIMILARITY).
FT METAL 169 169 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 206 AA; 23456 MW; B149F228DB49091E CRC64;

Query Match          2.5%; Score 6; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 202 SYSFWL 207
DB 175 SYSFWL 180

RESULT 108
JAG_BACHD          STANDARD;          PRT; 207 AA.
ID _JAG_BACHD
AC Q95CA6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE JAG protein homolog.
DE JAG OR BH4063.
GN Bacillus halodurans.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99356711; PubMed=10427704;
RA Takami H., Masui N., Nakasone K., Horikoshi K.;
RT "Replication origin region of the chromosome of alkaliphilic Bacillus halodurans C-125."
RL Biosci. Biotechnol. Biochem. 63:1134-1137(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S., Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- SIMILARITY: Contains 1 KH domain.
CC -!- SIMILARITY: Contains 1 R3H domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AB013492; BAB82681.1; -.
CC EMBL; AP001520; BAB07782.1; -.
CC PIR; G84157; G84157.
CC InterPro; IPR001374; R3H.
CC Pfam; PF01424; R3H; 1.
CC SMART; SM00393; R3H; 1.
CC RNA-binding; Complete proteome.
KW DOMAIN 91 141 KH.
FT DOMAIN 150 202
FT DOMAIN 150 202
SQ SEQUENCE 207 AA; 23130 MW; BE02AF774460A632 CRC64;

Query Match          2.5%; Score 6; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 59 QDLGTL 64
DB 98 QDLGTL 103

RESULT 109
NUGM PARTE          STANDARD;          PRT; 209 AA.
ID _NUGM PARTE
AC P15600;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```



DE NADH-ubiquinone oxidoreductase subunit 9 (EC 1.6.5.3) (EC 1.6.99.3)  
DE (Protein P1).  
GN NAD9.  
OS Parametium tetraurelia.  
OC Mitochondrion.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;  
OC Parametium.  
OX NCBI\_TaxID=5888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Stock 51;  
RX MEDLINE=90174913; PubMed=2308823;  
RA Pritchard A.E., Seilhamer J.J., Mahalingam R., Sable C.L.,  
RA Venuti S.E., Cummings D.J.,  
RT Nucleotide sequence of the mitochondrial genome of Parametium.;  
RL Nucleic Acids Res. 18:173-180(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87055241; PubMed=3023187;  
RA Pritchard A.E., Seilhamer J.J., Cummings D.J.,  
RT "Parametium mitochondrial DNA sequences and RNA transcripts for  
RT cytochrome oxidase subunit I, URFL, and three ORFs adjacent to the  
RT replication origin.";  
RT Gene 44:243-253(1986).  
CC -1- FUNCTION: Transfer of electrons from NADH to the respiratory  
CC chain. The immediate electron acceptor for the enzyme is believed  
CC to be ubiquinone (By similarity).  
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -1- SUBUNIT: Complex I is composed of about 30 different subunits.  
CC -1- SUBCELLULAR LOCATION: Matrix and cytoplasmic side of the  
CC mitochondrial inner membrane.  
CC -1- SIMILARITY: Belongs to the complex I 30 kDa subunit family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X15917; CAA34058.1; -;  
CC EMBL; M15275; AAA79259.1; -;  
CC F1R; S07725; S07725.  
CC InterPro; IPR001268; Complex1\_30K.  
CC Pfam; PF00329; Complex1\_30Kd; 1.  
CC ProDom; PD001581; Complex1\_30K; 1.  
CC PROSITE; PS00542; COMPLEX1\_30K; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
SQ SEQUENCE 209 AA; 23529 MW; D3D4477DC0BEP43 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 209;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 169 GQALAS 174  
Db 63 GQALAS 68  
  
RESULT 110  
GTS1 CAEEL STANDARD; PRT; 210 AA.  
AC Q9607; 210 AA.  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable glutathione S-transferase R07B1.4 (EC 2.5.1.18)  
DE (GST class-sigma).  
DE R07B1.4.  
GN R07B1.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Kershaw J.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Durbin R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Conjugation of reduced glutathione to a wide number of  
CC exogenous and endogenous hydrophobic electrophiles (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
CC -1- SIMILARITY: Belongs to the GST superfamily. Sigma family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z48621; CAA89541.2; -;  
CC HSP; P24472; IGUK.  
CC WormPep; R07B1.4; CE30562.  
CC GO; GO:0004364; F:glutathione transferase activity; NAS.  
CC GO; GO:0006803; P:glutathione conjugation reaction; NAS.  
CC GO; GO:0004046; GST\_Cterm.  
CC InterPro; IPR004045; GST\_Nterm.  
CC InterPro; IPR004045; GST\_C; 1.  
CC Pfam; PF00043; GST\_N; 1.  
CC Pfam; PF02798; GST\_C; 1.  
KW Hypothetical protein; Transferase.  
SQ SEQUENCE 210 AA; 23876 MW; 0E6010336B385D9B CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 210;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 26 SQTAT 31  
Db 62 SQTAT 67  
  
RESULT 111  
RS3 OCEIH STANDARD; PRT; 213 AA.  
AC P59182;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPSC OR OBO125.  
OS Oceanobacillus iheyensis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
OX NCBI\_TaxID=182710;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;  
RX MEDLINE=2220767; PubMed=12235376;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya  
RT Ridge and its unexpected adaptive capabilities to extreme  
RT environments.";  
RL Nucleic Acids Res. 30:3927-3935(2002).  
CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.

```
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP004593; BAC12081.1; -
CC HAMAP; MF 01309; -; 1.
CC InterPro; IPR004087; KH dom.
CC InterPro; IPR009019; KH prok.
CC InterPro; IPR004044; KH TYPE 2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC InterPro; IPR005704; S3_bact.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRfams; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS00823; KH TYPE 2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
CC FT DOMAIN 38 106 KH TYPE-2.
CC SQ SEQUENCE 213 AA; 23534 MW; 3543002B6C3B2934 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVPL 41
DB 169 EGTVPL 174
|||||

RESULT 112
SODF_AERPE
ID SODF_AERPE STANDARD; PRT; 214 AA.
AC Q9Y8H8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
GN SOD OR APE0741.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamano S.;
RT "A cambialistic SOD in a strictly aerobic hyperthermophilic archaeon,
RT Aeropyrum pernix."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kougi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
```

```
CC -!- COFACTOR: Binds 1 manganese or iron ion per subunit (By
CC similarity). Belongs to the iron/manganese superoxide dismutase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB012621; BAA76442.1; -
CC EMBL; AP000060; BAA79718.1; -
CC PIR; F72664; F72664.
CC HSP; Q08713; I306.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC PRINTS; PR01703; MNSODISMUTASE.
CC ProDom; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Iron; Complete proteome.
CC KW METAL 31 31. MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 79 79 MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 165 165 MANGANESE OR IRON (BY SIMILARITY).
CC FT METAL 169 169 MANGANESE OR IRON (BY SIMILARITY).
CC SQ SEQUENCE 214 AA; 24577 MW; 641122779485DFOA CRC64;

Query Match 2.5%; Score 6; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALRPIY 119
DB 18 ALRPIY 23
|||||

RESULT 113
FLAI_METJA
ID FLAI_METJA STANDARD; PRT; 217 AA.
AC Q58301;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B1 precursor.
GN FLAB1 OR MJ0891.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sulton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervilave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella.
CC -!- SIMILARITY: Belongs to the archaeal flagellin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; U67533; AAB98894.1; -;  
DR PIR; C64411; C64411.  
DR TIGR; MJ0891; -;  
DR InterPro; IPR002774; Arch\_flagellin.  
DR Pfam; PF01917; Arch\_flagellin; 1.  
KW Flagellum; Multigene family; Complete proteome.  
FT PROPEP 1 12  
FT CHAIN 13 217  
SQ SEQUENCE 217 AA; 22700 MW; 4374437380061565 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRG 7  
DB 7 LKGRG 12

## RESULT 114

FLA2\_METJA STANDARD; PRT; 217 AA.  
AC Q58302;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Flagellin B2 precursor.  
GN FLA2 OR MJ0892.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
ON NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kervilange A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073(1996).

CC -1- FUNCTION: Flagellin is the subunit protein which polymerizes to  
CC form the filaments of archaeal flagella.  
CC -1- SIMILARITY: Belongs to the archaeal flagellin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; U67533; AAB98895.1; -;  
DR PIR; D64411; D64411.  
DR TIGR; MJ0892; -;  
DR InterPro; IPR002774; Arch\_flagellin.  
DR Pfam; PF01917; Arch\_flagellin; 1.  
KW Flagellum; Multigene family; Complete proteome.  
FT PROPEP 1 12  
FT CHAIN 13 217  
SQ SEQUENCE 217 AA; 22700 MW; 4374437380061565 CRC64;

SQ SEQUENCE 217 AA; 22577 MW; 6E9D9435C243A82D CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKGRG 7  
DB 7 LKGRG 12

## RESULT 115

RS3\_BACST STANDARD; PRT; 217 AA.  
AC P23309;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 30S ribosomal protein S3 (BS2) (BS3/BS4).  
GN RSC  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
ON NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCA 1503;  
RX MEDLINE=91025633; PubMed=2222862;  
RA Kroemer W.J., Hatakeyama T., Kimura M.;  
RT "Nucleotide sequences of Bacillus stearothermophilus ribosomal  
protein genes: part of the ribosomal S10 operon.";  
RL Biol. Chem. Hoppe-Seyler 371:631-636(1990).  
RN [2]  
RP SEQUENCE OF 1-15.  
RC STRAIN=10;  
RX MEDLINE=75019590; PubMed=4607606;  
RA Yaguchi M., Matheson A.T., Visentin L.P.;  
RT "Prokaryotic ribosomal proteins: N-terminal sequence homologies and  
structural correspondence of 30 S ribosomal proteins from Escherichia  
coli and Bacillus stearothermophilus.";  
RL FEBS Lett. 46:296-300(1974).

CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -1- SIMILARITY: Contains 1 KH type-2 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; X54994; CAA38740.1; -;  
DR PIR; S10613; S10613.  
DR HAMAP; MF 01309; -;  
DR InterPro; IPR004087; KH dom.  
DR InterPro; IPR009019; KH\_prok.  
DR InterPro; IPR004044; KH\_TYPE\_2.  
DR InterPro; IPR001351; Ribosomal\_S3\_C.  
DR InterPro; IPR008282; Ribosomal\_S3\_N.  
DR InterPro; IPR005704; S3\_bact.  
DR Pfam; PF00013; KH; 1.  
DR Pfam; PF00189; Ribosomal\_S3\_C; 1.  
DR Pfam; PF00417; Ribosomal\_S3\_N; 1.  
DR SMART; SMO0322; KH; 1.  
DR TIGRFAVS; TIGR01009; rpsC\_bact; 1.  
DR PROSITE; PS00823; KH\_TYPE\_2; 1.  
DR PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
KW Ribosomal protein; RNA-binding; rRNA-binding.

```

FT INIT MET 0 0
FT DOMAIN 37 105 KH TYPE-2.
SQ SEQUENCE 217 AA; 24458 MW; 5A473C921758718C CRC64;

Query Match
Best Local Similarity 2.5%; Score 6; DB 1; Length 217;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 168 EGTVP 173

RESULT 116
RS3_BACSU STANDARD; PRT; 217 AA.
AC P21465;
DT 01-NOV-1991 (Rel. 18, Created)
DT 10-OCT-2003 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3 (BS) (BS2).
GN RPS3 OR BSU01220.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG38;
RX MEDLINE=98037503; PubMed=9371452;
RA Li X., Lindahl L., Sha Y., Zeng J.M.;
RT "Analysis of the Bacillus subtilis S10 ribosomal protein gene cluster
RT identifies two promoters that may be responsible for transcription of
RT the entire 15-kilobase S10-spc-alpha cluster.";
RL J. Bacteriol. 179:7046-7054(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124188; PubMed=8969501;
RA Yasumoto K., Liu H., Jeong S.M., Ohashi Y., Kakinuma S.,
RA Tanaka K., Kawamura F., Yoshikawa H., Takahashi H.;
RT "Sequence analysis of a 50 kb region between spoOH and rrmH on the
RT Bacillus subtilis chromosome.";
RL Microbiology 142:3039-3046(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Hada K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holteppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mollado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presceac E., Fujic P., Purrelle G., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler B., Wedler H., Weitzenecker T.,

```

```

RA Winters P., Wibat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 1-17.
RX MEDLINE=82219212; PubMed=6806564;
RA Higo K.-I., Otake E., Osawa S.;
RT "Purification and characterization of 30S ribosomal proteins from
RT Bacillus subtilis: correlation to Escherichia coli 30S proteins.";
RL Mol. Gen. Genet. 185:239-244(1982).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U43929; AAC45962.1; -.
DR EMBL; D50302; BAA08837.1; -.
DR EMBL; D50303; -. NOT ANNOTATED_CDS.
DR EMBL; Z99104; CAB1898.1; -.
DR PIR; B69699; B69699.
DR Subtilist; BG19005; rpsC.
DR HAMAP; MF_01309; -.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR009019; KH prot.
DR InterPro; IPR004044; KH TYPE 2.
DR InterPro; IPR003351; Ribosomal_S3_C.
DR InterPro; IPR008282; Ribosomal_S3_N.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00189; Ribosomal_S3_C; 1.
DR Pfam; PF00417; Ribosomal_S3_N; 1.
DR SMART; SM00332; KH; 1.
DR TIGRFAMs; TIGR01009; rpsC_bact; 1.
DR PROSITE; PS00823; KH TYPE 2; 1.
DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
DR Ribosomal protein; RNA-binding; Complete proteome.
FT INIT MET 0 105 KH TYPE-2.
FT DOMAIN 37 163 K -> H (IN REF. 4).
FT CONFLICT 163 163 P -> S (IN REF. 2).
SQ SEQUENCE 217 AA; 24201 MW; 607BACD05FE15AF4 CRC64;

Query Match
Best Local Similarity 2.5%; Score 6; DB 1; Length 217;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 168 EGTVP 173

RESULT 117
RS3_LACLA STANDARD; PRT; 217 AA.
ID RS3_LACLA
AC Q9CDW8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR LL2093.

```

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=IL1403;  
RM MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,  
Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403";  
RL Genome Res. 11:731-753(2001).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
in the 70S ribosome, positioning it for translation (By  
similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE006438; AK06191.1; -;  
CC PIR; E86886; E86886.  
CC HAMAP; MF\_01309; -; 1.  
CC InterPro; IPR004087; KH\_dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00013; KH; 1.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC SMART; SM00322; KH; 1.  
CC PROSITE; PS00823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
CC FT DOMAIN 38 106 KH TYPE-2.  
CC SQ SEQUENCE 217 AA; 24034 MW; CF5937A341B76BD4 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 36 EGTVP L 41  
DB 169 EGTVP L 174  
RESULT 118  
RS3\_STAEP STANDARD; PRT; 217 AA.  
AC Q99527;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPS3 OR SAV2244 OR SA2041 OR MW2163,  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain MW2).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=156878, 156879, 156620;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=MU50 / ATCC 700699, and N315;  
RX MEDLINE=213111952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus";  
RL Lancet 357:1225-1240(2001).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN=MW2;  
RX MEDLINE=22040717; PubMed=12043378;  
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
acquired MRSA";  
RL Lancet 359:1813-1827(2002).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
in the 70S ribosome, positioning it for translation (By  
similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AP003364; BAB58406.1; -;  
CC EMBL; AP003366; BAB43336.1; -;  
CC EMBL; AP004829; BAB96028.1; -;  
CC PIR; G90021; G90021.  
CC HAMAP; MF\_01309; -; 1.  
CC InterPro; IPR004087; KH\_dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00013; KH; 1.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC SMART; SM00322; KH; 1.  
CC TIGRfams; TIGR01009; rpsC\_bact; 1.  
CC PROSITE; PS00823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
CC FT DOMAIN 38 106 KH TYPE-2.  
CC SQ SEQUENCE 217 AA; 24100 MW; 174CA582EB0DF917 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 36 EGTVP L 41  
DB 169 EGTVP L 174  
RESULT 119  
RS3\_STAEP STANDARD; PRT; 217 AA.  
AC Q8CRG6;

```
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SEI818.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228.
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RX Phung Y.-Q., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Qin Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,
RA Yuan Z.-H.
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3 family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016750; AAO05459.1; .
CC DR HAMAP; MF 01309; .; 1.
CC DR InterPro; IPR004087; KH_dom.
CC DR InterPro; IPR009019; KH_prok.
CC DR InterPro; IPR004044; KH_TYPE_2.
CC DR InterPro; IPR001351; Ribosomal_S3_C.
CC DR InterPro; IPR008282; Ribosomal_S3_N.
CC DR InterPro; IPR005704; S3_bact.
CC DR Pfam; PF00013; KH; 1.
CC DR Pfam; PF00189; Ribosomal_S3_C; 1.
CC DR Pfam; PF00417; Ribosomal_S3_N; 1.
CC DR SMART; SM00322; KH; 1.
CC DR TIGRFAMS; TIGR01009; rpsC_bact; 1.
CC DR PROSITE; PS05823; KH_TYPE_2; 1.
CC DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC DR Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 38 106 KH TYPE-2.
SQ SEQUENCE 217 AA; 24189 MW; 42488FE049F9E751 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred.No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 169 EGTVP 174
|||||
RESULT 120
RS3_STRM STANDARD; PRT; 217 AA.
AC P59186;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR RS3 OR SMU.2021.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
```

```
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=2295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Perretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3 family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE015024; AAN59624.1; ALT_INIT.
CC DR HAMAP; MF 01309; .; 1.
CC DR InterPro; IPR004087; KH_dom.
CC DR InterPro; IPR009019; KH_prok.
CC DR InterPro; IPR004044; KH_TYPE_2.
CC DR InterPro; IPR001351; Ribosomal_S3_C.
CC DR InterPro; IPR008282; Ribosomal_S3_N.
CC DR InterPro; IPR005704; S3_bact.
CC DR Pfam; PF00013; KH; 1.
CC DR Pfam; PF00189; Ribosomal_S3_C; 1.
CC DR Pfam; PF00417; Ribosomal_S3_N; 1.
CC DR SMART; SM00322; KH; 1.
CC DR TIGRFAMS; TIGR01009; rpsC_bact; 1.
CC DR PROSITE; PS05823; KH_TYPE_2; 1.
CC DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC DR Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 29 97 KH TYPE-2.
SQ SEQUENCE 217 AA; 24122 MW; CC8EB247CF331538 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 217;
Best Local Similarity 100.0%; Pred.No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41
DB 169 EGTVP 174
|||||
RESULT 121
RS3_STRM STANDARD; PRT; 217 AA.
AC Q9WW37;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR SP0215 OR SFR0195.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6, SP#5, and ZR1;
```

Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0

KT Ribosomal protein; rRNA-binding; rRNA-binding; Complete proteome.  
PW 105 KH TYPE-2.  
FT DOMAIN 38  
SQ SEQUENCE 217 AA; 24046 MW; B54CA0663A248663 CRC64;

QY 36 EGTVPL 41  
|||||  
Db 169 EGTVPL 174

RESULT 122  
RS3\_STRPY STANDARD; PRT; 217 AA.  
AC Q9AIW8; P59185; Q8K8X2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE 30S ribosomal protein S3.  
DE GPSC OR SPY0056 OR SPYM3\_0046 OR SPB0048 OR SPYM18\_0057 OR GBS0064 OR  
GN SAG0064.  
CS Streptococcus pyogenes.  
OS Streptococcus pyogenes (serotype M3),  
OS Streptococcus pyogenes (serotype M18),  
OS Streptococcus agalactiae (serotype III), and  
OS Streptococcus agalactiae (serotype V).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus  
OC NCBI\_TAXID=1314, 198466, 186103, 216495, 216466;  
RN [1]

SEQUENCE FROM N.A.  
RP SPECIES=S.pyogenes; STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RC MEDLINE=21192864; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RN Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).

SEQUENCE FROM N.A.  
RP SPECIES=S.pyogenes; STRAIN=MGAS315 / Serotype M3;  
RX MEDLINE=22133808; PubMed=1212206;  
RA Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,  
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
RA Schlievert P.M., Musser J.M.;  
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
RT phage-encoded toxins, the high-virulence phenotype, and clone  
RT emergence.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).

SEQUENCE FROM N.A.  
RP SPECIES=S.pyogenes; STRAIN=SSI-1 / Serotype M3;  
RX MEDLINE=22683278; PubMed=12799345;  
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,  
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,  
RA Hayashi H., Hattori M., Hamada S.;  
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a  
RT large-scale genomic rearrangement in invasive strains and new insights  
RT into phage evolution.";  
RN Genome Res. 13:1042-1055 (2003).

SEQUENCE FROM N.A.  
RP SPECIES=S.pyogenes; STRAIN=MGAS9232 / Serotype M18;  
RX MEDLINE=21927593; PubMed=11917108;  
RA Smoot J.C., Barbian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
RA Sylva G.L., Sturdevant D.E., Rickles S.M., Porcella S.F.,  
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
RT "Genome sequence and comparative microarray analysis of serotype M18  
RT



RT group A Streptococcus strains associated with acute rheumatic fever  
RT outbreaks"; Acad. Sci. U.S.A. 99:4668-4673 (2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.agalactiae; STRAIN=NEM316 / Serotype III;  
RX MEDLINE=22242508; PubMed=12354221;  
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trier-Cuot P.,  
RA Kunst F.;  
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
RT invasive neonatal disease.";  
RL Mol. Microbiol. 45:1499-1513 (2002).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.agalactiae; STRAIN=2603 V/R / Serotype V;  
RX MEDLINE=2222998; PubMed=12200347;  
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
RA Iacobino E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,  
RA Rinaldi D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V Streptococcus agalactiae";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB006478; AAK3188.1; -  
CC EMBL; AE014137; AM78653.1; ALT INIT.  
CC EMBL; AP005141; BAC63143.1; ALT\_INIT.  
CC EMBL; AB009558; AL98882.1; -  
CC EMBL; AL766843; CAD45709.1; -  
CC EMBL; AE014194; AAW98972.1; -  
CC Sagalinst; GBS0064; -  
CC TIGR; SAG0064; -  
CC HAMAP; MF\_01309; -; 1.  
CC InterPro; IPR004087; KH\_dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00013; KH; 1.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC TIGRfams; TIGR01009; rpsC\_bact; 1.  
CC PROSITE; PS50823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
CC Ribosomal protein; rRNA-binding; Complete proteome.  
FT DOMAIN 38 106 KH TYPE-2  
RP SEQUENCE 217 AA; 24134 MW; C92C723219A4A7B CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVPL 41  
DB 169 EGTVPL 174  
RESULT 123  
RS3\_LISMO STANDARD; PRT; 218 AA.  
ID RS3\_LISMO  
AC Q927L3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPS3 OR LMO2626 OR LIN2775.  
GE Listeria monocytogenes, and  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639; 1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=L.monocytogenes, and L.innocua;  
RC STRAIN=EGD-e / Serovar 1/2a, and CLIP 11262 / Serovar 6a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Ertian K.-D., Ishii H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kierst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Matounam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Roland J.-A., Voss H., Wehland J., Cossart P.,  
RT "Comparative genomics of Listeria species";  
RL Science 294:849-852 (2001).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL591983; CAD0704.1; -  
CC EMBL; AL596173; CAC98001.1; -  
CC PIR; AB1403; AB1403.  
CC PIR; A11778; A11778.  
CC Listalist; LIN02775; -  
CC Listalist; LMO2626; -  
CC HAMAP; MF\_01309; -; 1.  
CC InterPro; IPR004087; KH\_dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00013; KH; 1.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC SMART; SM00322; KH; 1.  
CC TIGRfams; TIGR01009; rpsC\_bact; 1.  
CC PROSITE; PS50823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.

KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
FT DOMAIN 38 106 KH TYPE-2.  
SQ SEQUENCE 219 AA; 24294 MW; C7BC458B0855AA1D CRC64;

Query Match 2.5%; Score 6; DB 1; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41  
Db 169 EGTVP 174

RESULT 125  
RS3\_BACCR STANDARD; PRT; 219 AA.  
AC Q81J36;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPSC OR BC0137.  
OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=226900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608415; PubMed=12721630;  
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,  
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
RA Grechkin Y., Fusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
RA Overbeek R., Kyrpides N.,  
RT "Genome sequence of Bacillus cereus and comparative analysis with  
RT Bacillus anthracis."  
RL Nature 423:87-91(2003).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE016998; AAP07218.1; -.  
DR HAMAP; MF 01309; -; 1.  
DR InterPro; IPR004087; KH dom.  
DR InterPro; IPR009019; KH\_prok.  
DR InterPro; IPR004044; KH\_TYPE\_2.  
DR InterPro; IPR001351; Ribosomal\_S3\_C.  
DR InterPro; IPR008282; Ribosomal\_S3\_N.  
DR InterPro; IPR005704; S3\_bact.  
DR Pfam; PF00013; KH; 1.  
DR Pfam; PF00189; Ribosomal\_S3\_C; 1.  
DR Pfam; PF00417; Ribosomal\_S3\_N; 1.  
DR SMART; SM00322; KH; 1.  
DR TIGRFAMs; TIGR01009; tpSC\_bact; 1.  
DR PROSITE; PS08223; KH\_TYPE\_2; 1.  
DR PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
KW Ribosomal protein; RNA-binding; Complete proteome.  
FT DOMAIN 38 106 KH TYPE-2.  
SQ SEQUENCE 219 AA; 24294 MW; C7BC458B0855AA1D CRC64;

KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
FT DOMAIN 38 106 KH TYPE-2.  
SQ SEQUENCE 219 AA; 24542 MW; 4351A7E2CD75419A CRC64;

Query Match 2.5%; Score 6; DB 1; Length 218;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41  
Db 169 EGTVP 174

RESULT 124  
RS3\_BACAA STANDARD; PRT; 219 AA.  
AC Q81VS4;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPSC OR BA0116.  
OS Bacillus anthracis (strain Ames).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=198094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608414; PubMed=12721629;  
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzapple E.K., Ostad O.A., Helgason E., Ristone J., Wu M.,  
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Kouri H.M., Radune D.,  
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
RA Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria."  
RL Nature 423:81-86(2003).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE017024; AAP24170.1; -.  
DR TIGR; BA0116; -; 1.  
DR HAMAP; MF 01309; -; 1.  
DR InterPro; IPR004087; KH dom.  
DR InterPro; IPR009019; KH\_prok.  
DR InterPro; IPR004044; KH\_TYPE\_2.  
DR InterPro; IPR001351; Ribosomal\_S3\_C.  
DR InterPro; IPR008282; Ribosomal\_S3\_N.  
DR InterPro; IPR005704; S3\_bact.  
DR Pfam; PF00013; KH; 1.  
DR Pfam; PF00189; Ribosomal\_S3\_C; 1.  
DR Pfam; PF00417; Ribosomal\_S3\_N; 1.  
DR SMART; SM00322; KH; 1.  
DR TIGRFAMs; TIGR01009; tpSC\_bact; 1.  
DR PROSITE; PS08223; KH\_TYPE\_2; 1.

Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41  
| | | | |  
Db 169 EGTVP 174

## RESULT 126

RS3\_BACHD STANDARD; PRT; 219 AA.  
AC Q929K8; Q9JFY0;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S3.  
GN RPSC OR BH0140.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=99209008; PubMed=10192928;  
RA Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;  
RT "Sequence analysis of a 32-kb region including the major ribosomal  
protein gene clusters from alkaliphilic Bacillus sp. strain C-125.";  
RL Biosci. Biotechnol. Biochem. 63:452-455(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
Fuji F., Hirma C., Nakamura Y., Ogasawara N., Kuhara S.,  
Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
in the 70S ribosome, positioning it for translation (By  
similarity).  
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
with proteins S10 and S14 (By similarity).  
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -!- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB017508; BAA75277.1; -.  
CC EMBL; AP001507; BAB03859.1; -.  
CC PIR; T44389; T44389.  
CC HAMAP; MF\_01309; -; 1.  
CC InterPro; IPR004087; KH\_dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00013; KH; 1.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC SMART; SM00322; KH; 1.  
CC TIGRPFAMs; TIGR01009; rpsC\_bact; 1.  
CC PROSITE; PS50823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
DOMAIN 38 106 KH TYPE-2.

SQ SEQUENCE 219 AA; 24383 MW; 1F67970BBF67C97E CRC64;

Query Match 2.5%; Score 6; DB 1; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 EGTVP 41  
| | | | |  
Db 169 EGTVP 174

## RESULT 127

SODM\_HORSE STANDARD; PRT; 222 AA.  
AC Q9XS41;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1)  
(Mn-SOD).  
GN SOD2.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=99261591; PubMed=10331206;  
RX TISSUE=Testis;  
RA Ishida N., Katayama Y., Sato F., Hasegawa T., Mukoyama H.;  
RT "The cDNA sequences of equine antioxidant enzyme genes Cu/Zn-SOD and  
Mn-SOD, and these expressions in equine tissues.";  
RL J. Vet. Med. Sci. 61:291-294(1999).  
CC -!- FUNCTION: Destroys radicals which are normally produced within the  
cells and which are toxic to biological systems.  
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).  
CC -!- SUBUNIT: Homotrimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase  
family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB001693; BAA76922.1; -.  
CC HSPF; P04179; IABM.  
CC InterPro; IPR001189; SODismutase.  
CC Pfam; PF00081; scdfe; 1.  
CC Pfam; PF02777; scdfe; 1.  
CC PRINTS; PR01703; MNSODISMUTASE.  
CC ProDom; PD000475; SODismutase; 1.  
CC PROSITE; PS00088; SOD\_MN; 1.  
CC Oxidoreductase; Metal-binding; Manganese; Mitochondrion;  
KW Transit peptide.

TRANSIT 1 24 MITOCHONDRION (BY SIMILARITY).  
FT CHAIN 25 222 SUPEROXIDE DISMUTASE [MN].  
FT METAL 50 50 MANGANESE (BY SIMILARITY).  
FT METAL 98 98 MANGANESE (BY SIMILARITY).  
FT METAL 183 183 MANGANESE (BY SIMILARITY).  
FT METAL 187 187 MANGANESE (BY SIMILARITY).  
SQ SEQUENCE 222 AA; 24739 MW; 93A069491944E98C CRC64;

Query Match 2.5%; Score 6; DB 1; Length 222;  
Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYI 119

```
Db 37 ALEPYI 42
|||||
Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LGTLGS 66
DB 10 LGTLGS 15

RESULT 128
PGC2 HUMAN STANDARD; PRT; 223 AA.
ID PGC2 HUMAN STANDARD; PRT; 223 AA.
AC O15173;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Membrane associated progesterone receptor component 2 (Progesterone
DE membrane binding protein) (Steroid receptor protein D66).
GN PORWC2 OR PRBP OR DG6.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=9836853; PubMed=9705155;
RA Gerdes D., Wehling M., Leube B., Falkenstein E.;
RT "Cloning and tissue expression of two putative steroid membrane
RT receptors.";
RL Biol. Chem. 379:907-911 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Khatami S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Pamer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: IS A RECEPTOR FOR STEROIDS (POTENTIAL).
CC -1- SIMILARITY: Belongs to the MAPR family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ002030; CAA05152.1; -
DR EMBL; BC016892; AAH16892.1; -
DR Genew; HGNC:16089; PGRWC2.
DR MIM; 607735; -
DR GO; GO:0016021; C: integral to membrane; TAS.
DR GO; GO:0005496; F: steroid binding; TAS.
DR GO; GO:0003707; F: steroid hormone receptor activity; TAS.
DR InterPro; IPR001199; Cyt_B5.
DR Pfam; PF00173; heme_1; 1.
DR Receptor; Steroid-binding; Transmembrane; Microsome.
FT TRANSMEM 42 66
FT SEQUENCE 223 AA; 23818 MW; BE36229ED0FF3AD CRC64;

Query Match 2.5%; Score 6; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 LEKIIS 234
DB 171 LEKIIS 176

RESULT 130
LIPB MYCLE STANDARD; PRT; 235 AA.
ID LIPB MYCLE STANDARD; PRT; 235 AA.
```

```
AC 032961;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipote-protein ligase B (EC 6.-.-) (Lipoate biosynthesis protein
DE B).
DE
GN LIPE OR MLC859 OR MLCB22.19.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RX MEDLINE=21129732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churchworth T., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Involved in the attachment of lipoyl groups to proteins,
CC by creating an amide linkage that joins the free carboxyl group of
CC lipoyl acid to the epsilon-amino group of a specific lysine
CC residue in lipoylated proteins (By similarity).
CC
CC -!- PATHWAY: Lipote biosynthesis.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC -!- SIMILARITY: Belongs to the lipB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z98741; CAB11384.1; -.
CC EMBL; AL583920; CAC31240.1; -.
CC F01; T44894; T44894.
CC Lepronia; MLC859; -.
CC HAMAP; MF_00013; -.
CC InterPro; IPR004143; BPL LipA LipB.
CC InterPro; IPR005544; Lipotease B.
CC Pfam; PF03099; BPL LipA LipB; 1.
CC ProDom; PD006086; Lipotease; 1.
CC TIGRFAMs; TIGR00214; lipB; 1.
CC PROSITE; PS01313; LIPB; 1.
CC Ligase; Complete proteome.
CC
CC SEQUENCE 235 AA; 25002 MW; 353FA05D2EA4A67D CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 96 WLSTPA 101
DB 224 WLSTPA 229
|||||
RESULT 131
SOD1_PLEBO STANDARD; PRT; 248 AA.
AC P50058;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Mn] 1 precursor (EC 1.15.1.1).
```

```
GN SOD1.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxID=1184;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UTEX 485;
RX MEDLINE=95164530; PubMed=7860607;
RA Campbell W.S., Laudenbach D.E.;
RA "Characterization of four superoxide dismutase genes from a
RA filamentous cyanobacterium.";
RL J. Bacteriol. 177:964-972(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 manganese ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- INDUCTION: Constitutively expressed.
CC -!- SIMILARITY: Belongs to the iron/manganese superoxide dismutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U17609; AAA69950.1; -.
CC HSP; P00448; LVEW.
CC InterPro; IPR000437; Prok lipoprot_S.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe C; 1.
CC PRINTS; PR01703; MNSODISWTASE.
CC ProDom; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Metal-binding; Manganese; Multigene family; Signal.
CC
CC SIGNAL 1 41 POTENTIAL.
CC CHAIN 42 248 SUPEROXIDE DISMUTASE [MN] 1.
CC METAL 68 68 MANGANESE (BY SIMILARITY).
CC METAL 123 123 MANGANESE (BY SIMILARITY).
CC METAL 208 208 MANGANESE (BY SIMILARITY).
CC METAL 212 212 MANGANESE (BY SIMILARITY).
CC SEQUENCE 248 AA; 27955 MW; 85697F08623EA1D1 CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 248;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 ALEPYI 119
DB 55 ALEPYI 60
|||||
RESULT 132
TRYP_PLEPL STANDARD; PRT; 250 AA.
AC P35034;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Pleuronectes platessa (Plaice).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectidae; Pleuronectidae; Pleuronectes.
OX NCBI_TaxID=8262;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
```

```

RA Leaver M.J., George S.G.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X56744; CA940068.1; -
DR F01; S31384; S31384.
DR HSP; P00761; 1EPT.
DR MERO; S01151; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001134; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 22
FT ACT_SITE 23 250
FT CHAIN 23 250
FT ACT_SITE 62 62
FT ACT_SITE 106 106
FT ACT_SITE 203 203
FT DISULFID 29 163
FT DISULFID 47 63
FT DISULFID 133 236
FT DISULFID 140 209
FT DISULFID 174 188
FT DISULFID 199 223
FT SITE 197 197
FT SITE 250 250
FT SEQUENCE 250 AA; 27527 MW; 637DE96185CIABAA CRC64;

Query Match 2.5%; Score 6; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12
DB 201 GDSGSP 206

RESULT 133
ID_GRAA_MOUSE STANDARD; PRT; 260 AA.
AC P11032; P15118;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Granzyme A precursor (EC 3.4.21.78) (T cell-specific serine protease
DE 1) (TSP-1) (CTLA-3) (Fragmentin 1) (Autocrine thymic lymphoma
DE granzyme-like serine protease).
GN GZMA OR CTLA3 OR CTLA-3 OR WTSP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=92347841; PubMed=1639378;
RA Ebnet K., Kramer M.D., Simon M.M.;
RT "Organization of the gene encoding the mouse T-cell-specific serine

```

```

RT proteinase 'granzyme A'";
RL Genomics 13:502-508(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=89144553; PubMed=2976140;
RA Bogenberger J., Haas M.;
RT "cDNA clones from autocrine thymic lymphoma cells encode two
RT mitogenic proteins, a serine protease and a truncated T-cell receptor
RT beta-chain.";
RL Oncogene Res. 3:301-312(1988).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS HF1 AND HF2).
RX MEDLINE=93094270; PubMed=1460043;
RA Hershberger R.J., Gershenfeld H.K., Weissman I.L., Su L.;
RT "Genomic organization of the mouse, granzyme A gene. Two mRNAs encode
RT the same mature granzyme A with different leader peptides.";
RL J. Biol. Chem. 267:25488-25493(1992).
RN [4]
RP SEQUENCE OF 12-260 FROM N.A.
RX MEDLINE=86208119; PubMed=2422755;
RA Gershenfeld H.K., Weissman I.L.;
RT "Cloning of a cDNA for a T cell-specific serine protease from a
RT cytotoxic T lymphocyte.";
RL Science 232:854-858(1986).
RN [5]
RP SEQUENCE OF 29-48.
RX MEDLINE=87215912; PubMed=3555842;
RA Masson D., Tschopp J.;
RT "A family of serine esterases in lytic granules of cytolytic T
RT lymphocytes.";
RL Cell 49:679-685(1987).
RN [6]
RP SEQUENCE OF 29-53.
RX MEDLINE=87030960; PubMed=3533635;
RA Masson D., Zamaï M., Tschopp J.;
RT "Identification of granzyme A isolated from cytotoxic T-lymphocyte-
RT granules as one of the proteases encoded by CTL-specific genes.";
RL FEBS Lett. 208:84-88(1986).
RN [7]
RP SEQUENCE OF 29-46 FROM N.A.
RX MEDLINE=88255076; PubMed=3260181;
RA Simon H.G., Fruth U., Eckerskorn C., Lottspeich F., Kramer M.D.,
RA Nerz G., Simon M.M.;
RT "Induction of T cell serine proteinase 1 (TSP-1)-specific mRNA in
RT mouse T lymphocytes.";
RL Eur. J. Immunol. 18:855-861(1988).
RN [8]
RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=88272336; PubMed=3292396;
RA Jenne D.E., Tschopp J.;
RT "Granzymes, a family of serine proteases released from granules of
RT cytolytic T lymphocytes upon T cell receptor stimulation.";
RL Immunol. Rev. 103:53-71(1988).
CC -!- FUNCTION: This enzyme is necessary for target cell lysis in cell-
CC mediated immune responses. It cleaves after Lys or Arg. May be
CC involved in apoptosis.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,
CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,
CC Lys-|-Xaa >> Phe-|-Xaa in small molecule substrates.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Secreted; cytoplasmic granules of cytolytic
CC T-lymphocytes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=HF1;
CC IsoId=P11032-1; Sequence=Displayed;
CC Note=More abundant in lymphoid tissues than isoform HF2;
CC Name=HF2;
CC IsoId=P11032-2; Sequence=VSP_005377;
CC -!- TISSUE SPECIFICITY: Found in cytotoxic lymphocytes and in normal
CC lymphoid tissues such as thymus and spleen.
CC -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.

```





RP 3D-STRUCTURE MODELING.  
 RX MEDLINE=9184501; PubMed=3237717;  
 RA Murphy M.E.P., Moulton J., Bleackley R.C., Gershenfeld H.,  
 RA Weissman I.L., James M.N.G.;  
 RT "Comparative molecular model building of two serine proteinases from  
 RL cytototoxic T lymphocytes.";  
 RL Proteins 4:190-204(1988).  
 CC !- FUNCTION: This enzyme is necessary for target cell lysis in cell-  
 CC mediated immune responses. It cleaves after Lys or Arg. May be  
 CC involved in apoptosis.  
 CC !- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,  
 CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,  
 CC Lys-|-Xaa >> Phe-|-Xaa in small molecule substrates.  
 CC !- SUBUNIT: Homodimer; disulfide-linked.  
 CC !- SUBCELLULAR LOCATION: Secreted; cytoplasmic granules.  
 CC !- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M18737; AAA52647.1; -.  
 DR EMBL; BC015739; AAH15739.1; -.  
 DR EMBL; U40006; AAD00009.1; -.  
 DR EMBL; A31372; A31372  
 DR PIR; A31372; A31372  
 DR PDB; 1HF1; 15-OCT-94.  
 DR MEROPS; S01.135; -.  
 DR Genew; HGNC:4708; GZMA.  
 DR MIM; 140050; -.  
 DR GO; GO:0004277; F:Granzyme A activity; TAS.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001354; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1a.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRY SP; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;  
 KW Apoptosis; 3D-structure.  
 FT SIGNAL 1 26  
 FT PROPEP 27 28  
 FT CHAIN 29 262  
 FT ACT\_SITE 69 69  
 FT ACT\_SITE 114 114  
 FT ACT\_SITE 212 212  
 FT DISULFID 54 70  
 FT DISULFID 148 218  
 FT DISULFID 179 197  
 FT DISULFID 208 234  
 FT CARBOHYD 170 170  
 FT STRAND 30 30  
 FT STRAND 33 33  
 FT TURN 37 38  
 FT TURN 41 42  
 FT STRAND 43 48  
 FT TURN 49 51  
 FT STRAND 52 60  
 FT TURN 61 62  
 FT STRAND 63 66  
 FT TURN 68 69  
 FT STRAND 76 80  
 FT STRAND 84 84  
 FT TURN 85 86  
 FT TURN 90 91  
 FT STRAND 93 102  
 FT TURN 104 105  
 FT HELIX 108 110

FT TURN 113  
 FT STRAND 116  
 FT STRAND 120  
 FT TURN 127  
 FT TURN 129  
 FT STRAND 130  
 FT STRAND 134  
 FT TURN 138  
 FT TURN 144  
 FT STRAND 147  
 FT STRAND 155  
 FT TURN 158  
 FT STRAND 160  
 FT STRAND 165  
 FT STRAND 167  
 FT STRAND 173  
 FT HELIX 176  
 FT TURN 181  
 FT TURN 184  
 FT TURN 193  
 FT STRAND 195  
 FT TURN 201  
 FT STRAND 206  
 FT TURN 209  
 FT TURN 212  
 FT STRAND 215  
 FT TURN 219  
 FT STRAND 221  
 FT TURN 231  
 FT TURN 234  
 FT TURN 237  
 FT STRAND 241  
 FT TURN 246  
 FT HELIX 252  
 FT SEQUENCE 262 AA; 28966 MW; DA87363A0D92BAF4 CRC64;  
 SQ  
 Query Match 2.5%; Score 6; DB 1; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 GDSGSP 12  
 DB 210 GDSGSP 215  
 RESULT 135  
 ELNE HUMAN STANDARD; PRT; 267 AA.  
 AC P08246; P09649;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Leukocyte elastase precursor (SC 3.4.21.37) (Neutrophil elastase)  
 DE (PMN elastase) (Bone marrow serine protease) (Medullasin).  
 GN ELA2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCEI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89374820; PubMed=2775493;  
 RA Farley D., Travis J., Salvesen G.;  
 RT "The human neutrophil elastase gene. Analysis of the nucleotide  
 RL sequence reveals three distinct classes of repetitive DNA.";  
 RL Biol. Chem. Hoppe-Seyler 370:737-744(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88067782; PubMed=3479752;  
 RA Nakamura H., Okano K., Aoki Y., Shimizu H., Naruto M.;  
 RT "Nucleotide sequence of human bone marrow serine protease  
 RL (medullasin) gene.";  
 RL Nucleic Acids Res. 15:9601-9601(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=89008342; PubMed=2302087;  
RA Takahashi H., Nukiwa T., Yoshimura K., Quick C.D., States D.J.,  
RA Holmes M.D., Whang-Peng J., Knutson T., Crystal R.G.;  
RT "Structure of the human neutrophil elastase gene.";  
RL J. Biol. Chem. 263:14739-14747(1988).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90211319; PubMed=2322278;  
RX Okano K., Aoki Y., Shimizu H., Naruto M.;  
RA "Functional expression of human leukocyte elastase (HLE)/medullasin  
RT in eukaryotic cells.";  
RL Biochem. Biophys. Res. Commun. 167:1326-1332(1990).  
RN [5]  
RP SEQUENCE OF 30-267 FROM N.A.  
RX MEDLINE=89032918; PubMed=2822677;  
RX Okano K., Aoki Y., Sakurai T., Kajitani M., Kanai S., Shimazu T.,  
RA Shimizu H., Naruto M.;  
RT "Molecular cloning of complementary DNA for human medullasin: an  
RT inflammatory serine protease in bone marrow cells.";  
RL J. Biochem. 102:13-16(1987).  
RN [6]  
RP SEQUENCE OF 75-267 FROM N.A.  
RX MEDLINE=88115408; PubMed=3422232;  
RX Takahashi H., Nukiwa T., Bassett P., Crystal R.G.;  
RT "Myelomonocytic cell lineage expression of the neutrophil elastase  
RT gene.";  
RL J. Biol. Chem. 263:2543-2547(1988).  
RN [7]  
RP SEQUENCE OF 30-247  
RX MEDLINE=87175647; PubMed=3550808;  
RX Sinha S., Watorek W., Karr S., Giles J., Bode W., Travis J.;  
RT "Primary structure of human neutrophil elastase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2228-2232(1987).  
RN [8]  
RP SEQUENCE OF 262-267.  
RX MEDLINE=91315473; PubMed=1859409;  
RX Aoki Y., Hase T.;  
RT "The primary structure and elastolytic activity of medullasin (a  
RT serine protease of bone marrow).";  
RL Biochem. Biophys. Res. Commun. 178:501-506(1991).  
RN [9]  
RP PRELIMINARY SEQUENCE OF 30-103  
RA Travis J., Giles P.J., Porcelli L., Reilly C.F., Baugh R., Powers J.;  
RL (in) Protein degradation in health and disease, Ciba Foundation  
RL Symposium, pp.75:51-68, Excerpta Medica, Amsterdam and Oxford (1980).  
RN [10]  
RP SEQUENCE OF 30-49.  
RX MEDLINE=89315847; PubMed=2501794;  
RX Gabay J.E., Scott R.W., Campanelli D., Griffith J., Wilde C.,  
RA Marra M.N., Seeger M., Nathan C.F.;  
RT "Antibiotic proteins of human polymorphonuclear leukocytes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:5610-5614(1989).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (1.84 ANGSTROMS).  
RX MEDLINE=89098932; PubMed=2911584;  
RX Navia M.A., McKeever B.M., Springer J.P., Lin T.-Y., Williams H.R.,  
RA Fluder E.M., Dorn C.P., Hoogsteen K.;  
RT "Structure of human neutrophil elastase in complex with a peptide  
RT chloromethyl ketone inhibitor at 1.84-A resolution.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:7-11(1989).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=88271660; PubMed=3391280;  
RX Wei A.-Z., Mayr I., Bode W.;  
RT "The refined 2.3-A crystal structure of human leukocyte elastase in a  
RT complex with a valine chloromethyl ketone inhibitor.";  
RL FEBS Lett. 234:367-373(1988).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
RX MEDLINE=87053808; PubMed=3640709;  
RX Bode W., Wei A.-Z., Huber R., Meyer E., Travis J., Neumann S.;  
RA "X-ray crystal structure of the complex of human leukocyte elastase  
RT (PMN elastase) and the third domain of the turkey ovomucoid

RT inhibitor.";  
RL EMBO J. 5:2453-2458(1986).  
RN [14]  
RP VARIANTS CH VAL-32; PHE-177 AND GLN-191.  
RX MEDLINE=20047772; PubMed=10581030;  
RX Horwitz M., Benson K.F., Person R.E., Aprikyan A.G., Dale D.C.;  
RT "Mutations in ELA2, encoding neutrophil elastase, define a 21-day  
RT biological clock in cyclic haematopoiesis.";  
RL Nat. Genet. 23:433-436(1999).  
CC -|- FUNCTION: Medullasin modifies the functions of natural killer  
CC cells, monocytes and granulocytes.  
CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin.  
CC Preferential cleavage: Val-|-Xaa > Ala-|-Xaa.  
CC -|- TISSUE SPECIFICITY: Bone marrow cells.  
CC (CYCLIC NEUTROPENIA); AN AUTOSOMAL DOMINANT DISEASE IN WHICH  
CC BLOOD-CELL PRODUCTION FROM THE BONE MARROW OSCILLATES WITH 21-DAY  
CC PERIODICITY. CIRCULATING NEUTROPHILS VARY BETWEEN ALMOST NORMAL  
CC NUMBERS AND ZERO. DURING INTERVALS OF NEUTROPENIA, AFFECTED  
CC INDIVIDUALS ARE AT RISK FOR OPPORTUNISTIC INFECTION. MONOCYTES,  
CC PLATELETS, LYMPHOCYTES AND RETICULOCYTES ALSO CYCLE WITH THE SAME  
CC FREQUENCY.  
CC -|- SIMILARITY: Belongs to peptidase family S1. Elastase subfamily.  
CC -----  
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J03545; AAA52378.1; -;  
DR EMBL; Y00477; CAA68537.1; -;  
DR EMBL; X05875; CAA29299.1; -;  
DR EMBL; X05875; CAA29300.1; -;  
DR EMBL; M20203; AAA36359.1; -;  
DR EMBL; M20199; AAA36359.1; JOINED.  
DR EMBL; M20200; AAA36359.1; JOINED.  
DR EMBL; M20201; AAA36359.1; JOINED.  
DR EMBL; M34379; AAA36173.1; -;  
DR EMBL; D00187; BAA00128.1; -;  
DR PIR; A31976; ELHUL.  
DR PDB; 1HNE; 15-OCT-89.  
DR PDB; 1PPF; 31-JAN-94.  
DR PDB; 1B0F; 18-NOV-98.  
DR MEROPS; S01.131; -;  
DR Genew; HGNC:3309; ELA2.  
DR MIM; 130130; -;  
DR MIM; 162800; -;  
DR GO; GO:0004234; F:macrophage elastase activity; TAS.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SMC0020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR Hydrolase; Serine protease; Glycoprotein; Signal; 3D-structure;  
DR Disease mutation.  
KW SIGNAL 1 27 POTENTIAL.  
FT PROPEP 28 29  
FT CHAIN 30 267 LEUKOCYTE ELASTASE.  
FT ACT\_SITE 70 70 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 117 117 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM.  
FT DISULFID 55 71  
FT DISULFID 151 208  
FT DISULFID 181 187  
FT DISULFID 198 223

```
FT CARBOHYD 88      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124     N-LINKED (GLCNAC. . .)
FT CARBOHYD 173     N-LINKED (GLCNAC. . .)
FT VARIANT 32      G -> V (IN CH)
FT VARIANT 177     /FTid=VAR_009538.
FT VARIANT 191     V -> F (IN CH)
FT VARIANT 191     /FTid=VAR_009539.
FT CONFLICT 107    R -> Q (IN CH)
FT STRAND 31      /FTid=VAR_009540.
FT STRAND 31      N -> D (IN REF. 6).
FT STRAND 34      N -> D (IN REF. 6).
FT TURN 38
FT TURN 39
FT TURN 42
FT TURN 43
FT TURN 44
FT TURN 49
FT TURN 50
FT TURN 51
FT TURN 52
FT TURN 56
FT TURN 61
FT TURN 62
FT TURN 63
FT STRAND 64
FT STRAND 66
FT STRAND 67
FT STRAND 69
FT STRAND 71
FT TURN 72
FT TURN 74
FT TURN 77
FT TURN 79
FT STRAND 81
FT STRAND 84
FT STRAND 88
FT STRAND 89
FT TURN 90
FT TURN 94
FT TURN 95

Query Match      2.5%; Score 6; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDGSP 12
Db 200 GDGSP 205

RESULT 136
ZUPT_OCEIH STANDARD; PRT; 268 AA.
AC Q8ENQ1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc transporter zupt.
GN ZUPT OR OB2427.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments."
RT Nucleic Acids Res. 30:3927-3935 (2002).
RL NCBI_TaxID=182710;
CC -1- FUNCTION: Mediates zinc uptake. May also transport other divalent cations (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. Zupt subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP004601; BAC14383.1; -.
CC HAMAP; MF_00548; -. 1.
DR
```

```
DR InterPro; IPR003689; Zn_transpt_Zip.
DR Pfam; PF02535; Zip; 1.
KW Transport; Zinc transport; Transmembrane; Complete proteome.
FT TRANSMEM 4 26 Potential.
FT TRANSMEM 35 57 Potential.
FT TRANSMEM 72 94 Potential.
FT TRANSMEM 126 145 Potential.
FT TRANSMEM 155 177 Potential.
FT TRANSMEM 189 211 Potential.
FT TRANSMEM 215 237 Potential.
FT TRANSMEM 250 267 Potential.
SQ SEQUENCE 268 AA; 29073 MW; 393A3C339277E0C9 CRC64;

Query Match      2.5%; Score 6; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134
Db 156 AIAIAV 161

RESULT 137
YDGD_ECOLI STANDARD; PRT; 273 AA.
ID YDGD_ECOLI
AC P76176;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protease ydgd precursor (EC 3.4.21.-).
GN YDGD OR B1598.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474 (1997).
RL Science 277:1453-1474 (1997).
CC -1- SIMILARITY: Belongs to peptidase family S2B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000255; AAC74670.1; -.
CC PIR; H64915; H64915.
CC MEROPS; S01.260; -.
CC Ecogen; EG3925; ydgd.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR008256; Peptidase_S1B_V8.
CC Pfam; PF00089; tryptsin; 1.
CC PRINTS; PR00839; V8PROTEASE.
CC SMART; SM00020; Tryp_Src; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 PUTATIVE PROTEASE YDGD.
FT ACT_SITE 84 84 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 223 223 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR
```

SQ SEQUENCE 273 AA; 29277 MW; C8FCD018A59DBC62 CRC64;  
 Query Match 2.5%; Score 6; DB 1; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 GDSGSP 12  
 DB 221 GDSGSP 226  
 -----  
 RESULT 138  
 UPK\_STRPN STANDARD; PRT; 281 AA.  
 AC Q97SC8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative undecaprenol kinase (EC 2.7.1.66) (Bacitracin resistance protein).  
 GN UPK OR BACA OR SP0457 OR SPR0413.  
 OS Streptococcus pneumoniae, and  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313, 171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC BAA-334 / TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heideberg J., Deboy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Angioli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.; "Complete Genome sequence of a virulent isolate of Streptococcus pneumoniae.";  
 RT Science 293:498-506(2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC BAA-255 / R6;  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S., DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C., Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P., McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I., Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P., Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L., Glas J.I.; "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RT J. Bacteriol. 183:5709-5717(2001).  
 RL [3]  
 RN FUNCTION  
 RP STRAIN=0100993 / NCIMB 40794 / Serotype 3;  
 RX MEDLINE=20340958; PubMed=10878119;  
 RA Chalker A.F., Ingraham K.A., Lunsford R.D., Bryant A.P., Bryant J., Wallis N.G., Broskey J.P., Pearson S.C., Holmes D.J.; "The bcaA gene, which determines bacitracin susceptibility in Streptococcus pneumoniae and Staphylococcus aureus, is also required for virulence.";  
 RT Microbiology 146:1547-1553(2000).  
 RL  
 CC -!- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl phosphate. Confers resistance to bacitracin. Is also required for virulence.  
 CC -!- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition

of peptidoglycan synthesis by sequestering undecaprenyl diphosphate reducing the pool of lipid carrier available.  
 -!- SIMILARITY: Belongs to the upk family.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AE007357; AAK74617.1; -;  
 CC EMBL; AE008421; AAK99217.1; -;  
 CC PIR; E97923; E97923.  
 CC PIR; H95052; H95052.  
 CC TIGR; SP0457; -;  
 CC HAMAP; MF\_01006; -; 1.  
 CC InterPro; IPR003824; BACA.  
 CC Pfam; PF02673; BACA; 1.  
 CC Complete proteome.  
 KW Transferase; Kinase; Antibiotic resistance; Transmembrane;  
 FT TRANSMEM 10 32 POTENTIAL.  
 FT TRANSMEM 45 67 POTENTIAL.  
 FT TRANSMEM 111 133 POTENTIAL.  
 FT TRANSMEM 154 176 POTENTIAL.  
 FT TRANSMEM 191 213 POTENTIAL.  
 FT TRANSMEM 226 248 POTENTIAL.  
 FT TRANSMEM 263 280 POTENTIAL.  
 SQ SEQUENCE 281 AA; 31810 MW; 0575353B1BDF97C2 CRC64;  
 Query Match 2.5%; Score 6; DB 1; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 224 VKAGEL 229  
 DB 216 VKAGEL 221  
 -----  
 RESULT 139  
 AQP6\_HUMAN STANDARD; PRT; 282 AA.  
 AC Q13520;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Aquaporin 6 (Aquaporin-2 like) (hKID).  
 GN AQP6 OR AQP2L.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=97001157; PubMed=8812490;  
 RA Ma T., Yang B., Kuo W.L., Verkman A.S.; "cDNA cloning and gene structure of a novel water channel expressed exclusively in human kidney: evidence for a gene cluster of aquaporins at chromosome locus 12q13.";  
 RL Genomics 35:543-550(1996).  
 CC -!- FUNCTION: Forms a water-specific channel (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DOMAIN: Aquaporins contain two tandem repeats each containing three membrane-spanning domains and a pore-forming loop with the signature motif Asn-Pro-Ala (NPA).  
 CC -!- SIMILARITY: Belongs to the MIP/aquaporin (TC 1.A.8) family.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U48408; AAB41566.1; --  
CC Genew; HGNC:639; AQP6.  
CC MIM; 601383; --  
CC GO; GO:0005887; C: integral to plasma membrane; TAS.  
CC GO; GO:0015250; P: water channel activity; TAS.  
CC GO; GO:0007586; P: excretion; TAS.  
CC GO; GO:0006810; P: transport; TAS.  
CC InterPro; IPR000425; MIP.  
CC Pfam; PF00230; MIP; 1.  
CC PRINTS; PR00783; MINTRINSIC.  
CC ProDom; PD000295; MIP family; 1.  
CC TIGRfams; TIGR00861; MIP; 1.  
CC PROSITE; PS00221; MIP; 1.  
KW Transport; Repeat; Transmembrane.  
FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 31 48 POTENTIAL.  
FT DOMAIN 49 54 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 55 73 POTENTIAL.  
FT DOMAIN 74 99 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 100 121 POTENTIAL.  
FT DOMAIN 122 141 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 142 162 POTENTIAL.  
FT DOMAIN 163 168 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 169 188 POTENTIAL.  
FT DOMAIN 189 214 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 215 236 POTENTIAL.  
FT DOMAIN 237 282 CYTOPLASMIC (POTENTIAL).  
FT SITE 82 84 NPA 1.  
FT SITE 196 198 NPA 2.  
SQ SEQUENCE 282 AA; 29562 MW; 990BECB5D911ADD3 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 282;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 SGSPAT 14  
Db 169 SGSPAT 174  
  
RESULT 140  
ID TRUB\_FUSNN STANDARD; PRT; 287 AA.  
AC QRSY8;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55  
DE synthase) (Pse15 synthase) (Pseudouridylate synthase) (Uracil  
DE hydrolyase).  
GN TRUB OR FN0635.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
OC Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=21886394; PubMed=11899109;  
RA Kapatal V., Anderson I., Ivanova N., Resnik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusck G., Haselkorn R.,  
RA Fongstein M., Kyriades N., Overbeek R.,  
RT Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.  
RL J. Bacteriol. 184:2005-2018(2002).  
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from  
  
uracil-55 in the psi GC loop of transfer RNAs (By similarity).  
-!- CATALYTIC ACTIVITY: uracil + D-ribose 5-phosphate = pseudouridine  
-!- SIMILARITY: Belongs to the pseudouridine synthase trbB family.  
Subfamily 1.  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; A8010575; AAL94831.1; --  
HAMAP; MF\_01080; -- 1.  
InterPro; IPR004510; TrbB\_synth\_N.  
DR InterPro; IPR002501; TrbB\_N; 1.  
DR Pfam; PF01509; TrbB\_N; 1.  
DR TIGRfams; TIGR00431; TrbB; 1.  
KW tRNA processing; Lyase; Complete proteome.  
FT ACT SITE 38 38 BY SIMILARITY.  
SQ SEQUENCE 287 AA; 33092 MW; 5646C730F7BD6E23 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 287;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 58 GQDLGT 63  
Db 178 GQDLGT 183  
  
RESULT 141  
ID ZUPT\_CAMJE STANDARD; PRT; 291 AA.  
AC QPIN2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc transporter zupt.  
GN ZUPT OR CJ0263.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RC MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrall B.G.;  
RT The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.  
RL Nature 403:665-668(2000).  
CC -!- FUNCTION: Mediates zinc uptake. May also transport other divalent  
CC cations (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. Zupt  
CC subfamily.  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AL139074; CAB72731.1; --

DR PIR: F81444; F81444.  
DR HAMAP; MF\_00548; -; 1.  
DR InterPro; IPR003689; Zn\_transpt\_Zip.  
DR Pfam; PF02535; Zip; 1.  
DR KW Transport; Zinc transport; Transmembrane; Complete proteome.  
DR FT TRANSMEM 8 28 POTENTIAL.  
DR FT TRANSMEM 39 59 POTENTIAL.  
DR FT TRANSMEM 74 94 POTENTIAL.  
DR FT TRANSMEM 147 167 POTENTIAL.  
DR FT TRANSMEM 174 194 POTENTIAL.  
DR FT TRANSMEM 209 229 POTENTIAL.  
DR FT TRANSMEM 233 253 POTENTIAL.  
DR FT TRANSMEM 271 291 POTENTIAL.  
DR SQ SEQUENCE 291 AA; 31462 MW; 35A0E51E408E1CF2 CRC64;  
  
Query Match 2.5%; Score 6; DB 1; Length 291;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 129 AIAIAV 134  
DB 178 AIAIAV 183  
|||||  
  
RESULT 142  
EFTS\_XANCP STANDARD; PRT; 292 AA.  
ID EFTS\_XANCP STANDARD; PRT; 292 AA.  
AC Q8PAV3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Elongation factor Ts (EF-Ts).  
GN TSF OR XAC1421.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,  
RA Cicarella R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities."  
RL Nature 417:459-463(2002).  
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the  
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-  
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the EF-Ts family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
CC or send an email to license@isb-sib.ch).  
CC  
CC DR EMBL; AF012237; AM40672.1; ALT INIT.

```

DR HAVAP; MF_00050; -; 1.
DR InterPro; IPR001816; EF TS.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00889; EF TS; 1.
DR Pfam; PF00627; UBA; 1.
DR TIGRfams; TIGR00116; tsf; 1.
DR PROSITE; PS01126; EF TS_1; 1.
DR PROSITE; PS01127; EF TS_2; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
FT SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-
FT SITE 82 TU (BY SIMILARITY).
SQ SEQUENCE 292 AA; 31007 MW; A167E3F0AB576766 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 292;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234
DB 221 LEKIIIS 226
|||||

RESULT 144
EFTS_XYLFA STANDARD; PRT; 292 AA.
AC Q9PAD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Elongation factor Ts (EF-Ts).
DE TSF OR XF2579.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.K.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Pacinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.D., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.P.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.F., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Teuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the EF-Ts family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE004065; AAF95376.1; ALT_INIT.
DR HSSP; P02997; LEFU.
DR HAMAP; MF_00050; -; 1.
DR InterPro; IPR001816; UBA_domain.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00889; EF TS; 1.
DR Pfam; PF00627; UBA; 1.
DR TIGRfams; TIGR00116; tsf; 1.
DR PROSITE; PS01126; EF TS_1; 1.
DR PROSITE; PS01127; EF TS_2; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
FT SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-
FT SITE 82 TU (BY SIMILARITY).
SQ SEQUENCE 292 AA; 31235 MW; 433061318FACB8E0 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 292;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234
DB 221 LEKIIIS 226
|||||

RESULT 145
EFTS_XYLFT STANDARD; PRT; 292 AA.
AC Q87A70;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elongation factor Ts (EF-Ts).
DE TSF OR PD1959.
GN Xylella fastidiosa (strain Temeculal / ATCC 700964).
OS Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furian L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Camavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.L., Truffi D., Tsukumo F., Yanai G.M., Zatz L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the EF-Ts family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----

DR EMBL; AE012560; AAC29789.1; -;  
DR HANAP; MF\_00050; -; 1. -;  
DR InterPro; IPR001816; EF\_TS.  
DR InterPro; IPR000449; UBA\_domain.  
DR Pfam; PF00889; EF\_TS; 1.  
DR Pfam; PF00627; UBA; 1.  
DR PROSITE; PS01126; EF\_TS 1; 1.  
DR PROSITE; PS01127; EF\_TS 2; 1.  
KW Elongation factor; Protein biosynthesis; Complete proteome.  
FT SITE 79 INVOLVED IN MG++ ION DISLOCATION FROM EF-  
82 TU (BY SIMILARITY).  
SQ SEQUENCE 292 AA; 31347 MW; 7C2BBA34D56B6E0F CRC64;

Query Match 2.5%; Score 6; DB 1; Length 292;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LEKIIIS 234  
DB 221 LEKIIIS 226

RESULT 146  
VGLG\_HRSV5 STANDARD; PRT; 298 AA.  
AC P27024;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Major surface glycoprotein G (Attachment glycoprotein G).  
GN G.

OS Human respiratory syncytial virus (strain rsb6190).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
OX NCBI\_TaxID=11255;  
RN [1]\_TaxID=11255;  
RX SEQUENCE FROM N.A. PubMed=1895054;  
RP MEDLINE=91374005; PubMed=1895054;  
RA Cane P.A., Matthews D.A., Pringle C.R.;  
RT "Identification of variable domains of the attachment (G) protein of  
RT subgroup A respiratory syncytial viruses.";  
RL J. Gen. Virol. 72:2091-2096(1991).  
CC -!- FUNCTION: Unlike the other paramyxovirus attachment proteins, the  
CC respiratory syncytial virus G protein lacks both neuraminidase and  
CC hemagglutinating activities.  
CC -!- SUBCELLULAR LOCATION: Expressed on the surface of the infected  
CC cells and incorporated in the membrane of the virions.  
CC -!- PTM: May carry 40-80 separate O-linked carbohydrate chains  
CC distributed among the 91 serine and threonine residues.  
DR PIR; JQ1207; JQ1207.  
DR InterPro; IPR000925; Glycoprot G.  
DR Pfam; PF00802; Glycoprotein G; 1.  
KW Transmembrane, Glycoprotein.

FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 38 66 POTENTIAL.  
FT DOMAIN 67 298 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 298 AA; 32769 MW; 4D74E854D34D7BA5 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 298;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SQTTAI 31  
DB 109 SQTTAI 114  
RESULT 147  
GFR4\_HUMAN STANDARD; PRT; 299 AA.  
ID GFR4\_HUMAN  
AC Q9GZ77; Q9H191; Q9H192;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE GDNF family receptor alpha 4 precursor (GFR-alpha 4) (GFRalpha4)  
DE (Persephin receptor).  
GN GFR4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A. (ISOFORMS GFRALPHA4A; GFRALPHA4B AND GFRALPHA4C),  
RP AND GPI-ANCHOR.  
RC TISSUE=Thyroid;  
RX MEDLINE=21153758; PubMed=11116144;  
RA Lindahl M., Poteryaev D., Yu L., Arumae U., Timmusk T., Bongarzoni I.,  
RA Aiello A., Pierotti M.A., Airaksinen M.S., Saarma M.;  
RT "Human glial cell line-derived neurotrophic factor receptor alpha4 is  
RT the receptor for persephin and is predominantly expressed in normal  
RT and malignant thyroid medullary cells.";  
RL J. Biol. Chem. 276:9344-9351(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM GFRALPHA4A).  
RA Zhou B., Levinson B., Gitschier J.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. PubMed=11780052;  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Brown A.J.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Eutier A.P., Garder C., Carter N.P., Clee C.M.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammon S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Leharvaisho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Showkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
CC -!- FUNCTION: Receptor for persephin. Mediates the GDNF-induced  
CC autophosphorylation and activation of the RET receptor. May be  
CC important in C-cell development and, in the postnatal development  
CC of the adrenal medulla.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (isoforms GFRalpha4a and GFRalpha4b). Secreted (isoform  
CC GFRalpha4c).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;

```

CC CC Comment-Additional isoforms seem to exist;
CC CC Name=GFRalpha4b;
CC CC IsoId=Q9GZ27-1; Sequence=Displayed;
CC CC Name=GFRalpha4a;
CC CC IsoId=Q9GZ27-2; Sequence=VSP_007223;
CC CC Name=GFRalpha4c;
CC CC IsoId=Q9GZ27-3; Sequence=VSP_007224; VSP_007225;
CC CC TISSUE SPECIFICITY: Predominantly expressed in the adult thyroid
CC CC gland. Low levels also found in fetal adrenal and thyroid glands.
CC CC -!- SIMILARITY: Belongs to the GDNFR family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AJ291673; CAC19690.1; -
CC CC EMBL; AJ291674; CAC19691.1; -
CC CC EMBL; AJ291675; CAC19692.1; -
CC CC EMBL; AF253318; AAG25925.1; -
CC CC EMBL; AL356755; CAC16508.2; -
CC CC GenBank; HGNC:13821; GFR4.
CC CC InterPro; IPR003438; GDNF_receptor.
CC CC Pfam; PF02351; GDNF; 1.
CC CC PRINTS; PR01316; GDNFRECEPTOR.
CC CC Receptor; Glycoprotein; GPI-anchor; Membrane; Signal;
KW Alternative splicing; Lipoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 278 GDNF FAMILY RECEPTOR ALPHA 4.
FT PROPEP 279 299 REMOVED IN NATURE FORM (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 278 278 GPI-anchor amidated glycine (POTENTIAL).
FT VARSPLIC 132 197 CARAAGPWRGNGSLSPAPRPAAQSPGLSLGVHPSAQ
FT FT RPRRLPAGPG -> PRLAFOVSCTPAPSPADGGLLQOGR
FT FT AFSAPDGLDQARCLRAYAGLV (in isoform
FT FT GFRalpha4a).
FT FT /FTID=VSP_007223.
FT FT CARAAGPWRGNGRGLSPAPRPAAQSPGLSLGVHPSAQ
FT FT RPRRLPAGPG -> PRLAFOVSCTPAPSPADGGLLQOGR
FT FT CLRAYAGLVSPQAPPSPPLTTWT (in isoform
FT FT GFRalpha4c).
FT FT /FTID=VSP_007224.
FT FT Missing (in isoform GFRalpha4c).
FT FT VARSPLIC 183 299 /FTID=VSP_007225.
FT FT SEQUENCE 299 AA; 31669 MW; 8443B832FF10801 CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 299;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 TGRALE 116
Db 277 TGRALE 282
RESULT 148
CAFR_YERPE
ID CAFR_YERPE STANDARD; PRT; 301 AA.
AC P26950; Q93376.
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE F1 operon positive regulatory protein.
GN CAFIR OR YPM1.81C OR Y5193 OR Y1097.
OS Yersinia pestis.
OC Plasmid pMT1 (pMT-1), and Plasmid pFra.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC PLASMID=pFra;
RX MEDLINE=91323540; PubMed=1677900;
RA Galyov E.E., Karlishhev A.V., Chernovskaya T.V., Dolgikh D.A.,
RA Smirnov O.Y., Volkovoy K.I., Abramov V.M., Zav'yalov V.P.;
RT "Expression of the envelope antigen F1 of Yersinia pestis is mediated
RT by the product of cafM gene having homology with the chaperone
RT protein PapD of Escherichia coli.";
RL FEBS Lett. 286:79-82(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID=pFra;
RX MEDLINE=92339520; PubMed=1633857;
RA Karlyshev A.V., Galyov E.E., Abramov V.M., Zav'yalov V.P.;
RT "CafIR gene and its role in the regulation of capsule formation of Y.
RT pestis.";
RL FEBS Lett. 305:37-40(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen.";
RL Infect. Immun. 66:5731-5742(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCreedy P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Yersinia
RT pestis.";
RL J. Bacteriol. 180:5192-5202(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-32 / Biovar Orientalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaiha M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brookes K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- FUNCTION: POSITIVE REGULATOR OF F1 OPERON EXPRESSION.
CC -!- SIMILARITY: BELONGS TO THE ARAC/XYLIS FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X61996; CAA43969.1; -
CC CC EMBL; AF074611; AAC82755.1; ALT_INIT.
CC CC EMBL; AF053947; AAC13221.1; -
CC CC EMBL; AL117211; CAB55263.1; ALT_INIT.
CC CC PIR; S19097; S19097.
CC CC HSSP; P27246; 1BL0.
CC CC InterPro; IPR000005; HTHARAC.
CC CC Pfam; PF00165; HTH_AraC; 2.
CC CC PRINTS; PR00032; HTHARAC.
CC CC SMART; SM00342; HTH_ARAC; 1.
CC CC PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
CC CC PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
CC CC Transcription regulation; Activator; DNA-binding; Plasmid;
KW

```

```
KW Complete proteome.
FT DNA_BIND 25 44 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 124 124 E -> V (IN REF. 1).
FT CONFLICT 135 135 E -> V (IN REF. 1).
SQ SEQUENCE 301 AA; 36053 MW; E1EB5AD5C4CD43C0 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 301;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 LYSGPS 46
DB 30 LYSGPS 35

RESULT 149
Y548 STAEF STANDARD; PRT; 302 AA.
AC Q5CTE3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0042 protein SE0548.
GN SE0548.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qian Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -1- SIMILARITY: Belongs to the UPF0042 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016745; AAC004145.1; -.
CC HAMAP; MF 00636; -.
CC InterPro; IPR005337; UPF0042.
CC Pfam; PF03668; ATP_bind2; 1.
CC KW Hypothetical protein; ATP-binding; Complete proteome.
CC NP_BIND 18 25 ATP (POTENTIAL)
SQ SEQUENCE 302 AA; 34751 MW; 4516D8DA894DA5F CRC64;

Query Match 2.5%; Score 6; DB 1; Length 302;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKISR 235
DB 104 EKISR 109

RESULT 150
Y367 RICPR STANDARD; PRT; 303 AA.
AC Q9ZDG2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP367.
GN RP367.
```

```
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=90039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ235271; CAA14826.1; -.
CC PIR; H71693; H71693.
CC InterPro; IPR007487; DUF534.
CC Pfam; PF04392; DUF534; 1.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 303 AA; 33569 MW; D7703DEB5F29440 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 303;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 PSFLFV 50
DB 11 PSFLFV 16

Search completed: April 5, 2004, 07:37:55
Job time : 24 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:30:34 ; Search time 45 Seconds  
(without alignments)  
1710.810 Million cell updates/sec

Title: US-10-032-221B-10  
Perfect score: 244  
Sequence: 1 GLKRGKDSGSPATWTRGF.....KAGELEKIIISRCQVCMKKRH 244

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_todent.\*
- 12: sp\_virus.\*
- 13: sp Vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	100.0	245	4 QNYC4	QNYC4 homo sapien
2	117	48.0	212	6 Q28512	Q28512 macaca mula
3	54	22.1	203	6 Q28682	Q28682 cryctologus
4	39	16.0	112	11 Q8CCD6	Q8CCD6 mus musculus
5	39	16.0	230	11 Q63122	Q63122 ratrus norv
6	39	16.0	246	11 Q61435	Q61435 mus musculus
7	39	16.0	1669	11 Q9QZ50	Q9QZ50 mus musculus
8	38	15.6	161	11 Q61430	Q61430 mus musculus
9	36	14.8	52	4 Q9UDK9	Q9UDK9 homo sapien
10	36	14.8	56	4 Q9UDL0	Q9UDL0 homo sapien
11	36	14.8	97	4 Q8UDL1	Q8UDL1 homo sapien
12	35	14.3	212	6 Q28567	Q28567 ovis aries
13	25	10.2	210	6 Q28273	Q28273 canis famli
14	24	9.8	203	6 Q29032	Q29032 sus scrofa
15	18	7.4	29	4 Q9UDJ9	Q9UDJ9 homo sapien
16	17	7.0	179	11 P70165	P70165 mus musculus

17	7.0	225	6	Q28271	Q28271 canis famli
18	7.0	226	11	Q391Q8	Q391Q8 mus musculus
19	7.0	229	4	Q9NF88	Q9NF88 homo sapien
20	7.0	229	4	Q9NYCS	Q9NYCS homo sapien
21	7.0	253	11	Q61436	Q61436 mus musculus
22	7.0	585	11	Q8OV57	Q8OV57 mus musculus
23	7.0	799	11	Q8BNS7	Q8BNS7 mus musculus
24	7.0	886	4	Q8NUB7	Q8NUB7 homo sapien
25	7.0	979	13	Q319K3	Q319K3 gallus gall
26	7.0	1075	4	Q8EX41	Q8EX41 homo sapien
27	7.0	1621	4	Q8H4R9	Q8H4R9 homo sapien
28	7.0	1684	6	Q8HYC1	Q8HYC1 canis famli
29	7.0	1688	6	Q866Z2	Q866Z2 canis famli
30	7.0	1691	11	Q9ESQ2	Q9ESQ2 mus musculus
31	4.9	1752	5	Q07265	Q07265 strongyloce
32	4.5	713	5	Q9GV24	Q9GV24 sarcophaga
33	4.5	1761	5	Q18407	Q18407 drosophila
34	4.5	1940	5	Q9VMV5	Q9VMV5 drosophila
35	4.1	673	4	Q14052	Q14052 homo sapien
36	4.1	1024	5	Q8T7S4	Q8T7S4 anopheles g
37	4.1	1747	5	Q26640	Q26640 strongyloce
38	3.7	59	11	Q8BPI9	Q8BPI9 mus musculus
39	3.7	312	11	Q84457	Q84457 mus musculus
40	3.7	358	11	Q91V13	Q91V13 mus musculus
41	3.7	1682	11	Q9QZK9	Q9QZK9 mus musculus
42	3.3	74	2	Q48996	Q48996 mycoplasma
43	3.3	179	5	Q8T2W5	Q8T2W5 trypanosoma
44	3.3	304	16	Q81T03	Q81T03 bacillus an
45	3.3	337	11	Q8BXT8	Q8BXT8 mus musculus
46	3.3	337	11	Q8BGP6	Q8BGP6 mus musculus
47	3.3	452	11	Q921S0	Q921S0 mus musculus
48	3.3	486	11	Q8BN95	Q8BN95 mus musculus
49	3.3	523	17	Q8THZ6	Q8THZ6 methanosarc
50	3.3	541	11	Q8C7W7	Q8C7W7 mus musculus
51	3.3	546	11	Q99K97	Q99K97 mus musculus
52	3.3	1477	3	Q8C250	Q8C250 neurospora
53	3.3	1691	11	Q9ESQ1	Q9ESQ1 mus musculus
54	3.3	1723	5	Q9QCB1	Q9QCB1 hydra atten
55	3.3	1779	5	Q9VMV4	Q9VMV4 drosophila
56	2.9	49	2	Q9R4U1	Q9R4U1 azotobacter
57	2.9	68	16	Q8XW59	Q8XW59 ralstonia s
58	2.9	69	2	Q84ER2	Q84ER2 ralstonia o
59	2.9	109	12	Q67050	Q67050 influenzavi
60	2.9	109	12	Q67053	Q67053 influenzavi
61	2.9	109	12	Q67051	Q67051 influenzavi
62	2.9	109	12	Q67052	Q67052 influenzavi
63	2.9	116	16	Q89PD4	Q89PD4 bradyrhizob
64	2.9	118	16	Q8EKD6	Q8EKD6 shewanella
65	2.9	122	17	Q9YEN1	Q9YEN1 aeropyrum p
66	2.9	125	2	Q68196	Q68196 klebsiella
67	2.9	128	17	Q96YI1	Q96YI1 sulfolobus
68	2.9	157	8	Q9MJR3	Q9MJR3 taenia pisi
69	2.9	168	17	Q9YDQ8	Q9YDQ8 aeropyrum p
70	2.9	171	5	Q97232	Q97232 plasmodium
71	2.9	173	8	Q8HKQ0	Q8HKQ0 apasma min
72	2.9	174	16	Q82WJ0	Q82WJ0 nitrosomonas
73	2.9	175	5	Q8SVV5	Q8SVV5 encephalito
74	2.9	182	4	Q8ND70	Q8ND70 homo sapien
75	2.9	187	17	Q97ZW3	Q97ZW3 sulfolobus
76	2.9	189	2	Q59673	Q59673 propionibac
77	2.9	193	2	Q9ZGN1	Q9ZGN1 azotobacter
78	2.9	199	12	Q8TR59	Q8TR59 human reapi
79	2.9	200	16	Q8A303	Q8A303 caulobacter
80	2.9	202	6	Q28272	Q28272 canis famli
81	2.9	202	16	Q83HP6	Q83HP6 tropheryma
82	2.9	202	16	Q83GI4	Q83GI4 tropheryma
83	2.9	205	6	Q28274	Q28274 canis famli
84	2.9	206	3	Q96UT6	Q96UT6 candida alb
85	2.9	208	6	Q29468	Q29468 canis famli
86	2.9	211	17	Q96I84	Q96I84 sulfolobus
87	2.9	216	2	P95356	P95356 neisseria g
88	2.9	216	16	Q9K1R3	Q9K1R3 neisseria m
89	2.9	216	16	Q9JW11	Q9JW11 neisseria m

90	7	2.9	220	5	Q8MU16	Q8mu16 trichinella	163	7	2.9	471	12	Q9E311	Q9e311 influenza a
91	7	2.9	233	4	Q8IXG7	Q8ixg7 homo sapien	164	7	2.9	471	12	Q9E313	Q9e313 influenza a
92	7	2.9	236	16	Q7VAV7	Q7vav7 prochloroc	165	7	2.9	472	2	Q8GBR8	Q8gb8 treponema m
93	7	2.9	237	16	Q9KBB8	Q9kbb8 bacillus ha	166	7	2.9	472	12	Q9E310	Q9e310 influenza a
94	7	2.9	239	16	Q8FAX2	Q8fax2 leptospira	167	7	2.9	473	12	Q9E315	Q9e315 influenza a
95	7	2.9	241	5	Q8ML71	Q8ml71 drosophila	168	7	2.9	477	4	Q9BUD0	Q9bud0 homo sapien
96	7	2.9	244	4	Q8N2M3	Q8n2m3 homo sapien	169	7	2.9	478	16	Q9WZ22	Q9wz22 thermotoga
97	7	2.9	245	16	Q8G726	Q8g726 bifidobacte	170	7	2.9	494	4	Q8H0M4	Q8h0m4 homo sapien
98	7	2.9	247	16	Q8EXH8	Q8exh8 salmonella	171	7	2.9	503	12	Q8JL66	Q8jl66 ectromelia
99	7	2.9	252	16	Q87BW1	Q87bw1 xyfella fas	172	7	2.9	503	12	Q72739	Q72739 cowpox viru
100	7	2.9	254	16	Q81YK7	Q81yk7 bacillus an	173	7	2.9	511	10	Q8LRA0	Q8lra0 cryza sativ
101	7	2.9	254	16	Q81AT7	Q81at7 bacillus ce	174	7	2.9	523	4	Q99993	Q99993 homo sapien
102	7	2.9	255	10	Q9ZWM8	Q9zwm8 cryza sativ	175	7	2.9	548	12	Q919W4	Q919w4 influenza a
103	7	2.9	257	16	Q9PB46	Q9pb46 xyfella fas	176	7	2.9	548	12	Q919W5	Q919w5 influenza a
104	7	2.9	259	8	Q85M64	Q85m64 toxoplasma	177	7	2.9	548	12	Q919W3	Q919w3 influenza a
105	7	2.9	259	10	Q9SNQ0	Q9snq0 cryza sativ	178	7	2.9	549	6	Q86423	Q86423 bos taurus
106	7	2.9	264	16	Q34535	Q34535 bacillus su	179	7	2.9	550	12	Q8QLT9	Q8qlt9 influenza a
107	7	2.9	267	16	Q92C22	Q92c22 listeria in	180	7	2.9	550	12	Q98VJ6	Q98vu6 influenza a
108	7	2.9	267	16	Q8Y880	Q8y880 listeria mo	181	7	2.9	550	12	Q82499	Q82499 influenzavi
109	7	2.9	267	16	Q8ER45	Q8er45 oceanobacil	182	7	2.9	550	12	Q82498	Q82498 influenzavi
110	7	2.9	274	16	Q7V2M5	Q7v2m5 prochloroco	183	7	2.9	550	12	Q82753	Q82753 influenzavi
111	7	2.9	282	10	Q9LWS3	Q9lws3 cryza sativ	184	7	2.9	565	12	Q67103	Q67103 influenza a
112	7	2.9	285	17	Q8QW83	Q8qw83 methanosarc	185	7	2.9	565	12	Q67107	Q67107 influenza a
113	7	2.9	297	16	Q8F0K1	Q8f0k1 leptospira	186	7	2.9	565	12	Q67104	Q67104 influenza a
114	7	2.9	304	16	Q8GCA3	Q8gca3 pseudomonas	187	7	2.9	565	12	Q67102	Q67102 influenza a
115	7	2.9	313	11	Q8C7L9	Q8c7l9 mus musculu	188	7	2.9	565	12	Q91E65	Q91e65 influenza v
116	7	2.9	313	17	Q8Z2Y2	Q8z2y2 pyrobaculum	189	7	2.9	565	12	Q91E65	Q91e65 influenza v
117	7	2.9	316	16	Q8L5H8	Q8l5h8 rhizobium l	190	7	2.9	565	12	Q82559	Q82559 influenza a
118	7	2.9	331	16	Q88B27	Q88b27 pseudomonas	191	7	2.9	565	12	Q67105	Q67105 influenza a
119	7	2.9	338	16	Q92766	Q92t66 rhizobium m	192	7	2.9	565	12	Q82793	Q82793 influenzavi
120	7	2.9	343	10	Q9SAC7	Q9sac7 arabidopsis	193	7	2.9	565	12	Q91E71	Q91e71 influenzavi
121	7	2.9	354	16	Q915Y7	Q915y7 pseudomonas	194	7	2.9	565	12	Q82792	Q82792 influenzavi
122	7	2.9	365	12	Q56962	Q56962 influenza a	195	7	2.9	565	12	Q67106	Q67106 influenza a
123	7	2.9	365	12	Q56961	Q56961 influenza a	196	7	2.9	565	12	Q66752	Q66752 equine infl
124	7	2.9	371	12	Q3DL24	Q3dl24 influenza a	197	7	2.9	565	12	Q86639	Q86639 equine infl
125	7	2.9	372	4	Q9NT4	Q9nt4 homo sapien	198	7	2.9	566	12	Q98052	Q98052 influenzavi
126	7	2.9	372	4	Q8N3J5	Q8n3j5 homo sapien	199	7	2.9	566	12	Q37160	Q37160 influenzavi
127	7	2.9	372	4	Q8TUZ7	Q8tuz7 homo sapien	200	7	2.9	566	12	Q82496	Q82496 influenzavi
128	7	2.9	382	10	Q8H666	Q8h666 cryza sativ							
129	7	2.9	384	12	Q8JK63	Q8jk63 influenza a							
130	7	2.9	389	10	Q7XSV3	Q7xsv3 cryza sativ							
131	7	2.9	398	16	Q89C31	Q89c31 bradyrhizob							
132	7	2.9	403	4	Q9NWF7	Q9nwf7 homo sapien							
133	7	2.9	407	16	Q8EG37	Q8eg37 shewanella							
134	7	2.9	409	12	Q9Q0L5	Q9q0l5 influenza a							
135	7	2.9	412	2	Q91S15	Q91s15 burkholderi							
136	7	2.9	412	2	Q69122	Q69122 burkholderi							
137	7	2.9	416	2	Q8KN89	Q8kn89 pseudomonas							
138	7	2.9	416	2	Q8KN80	Q8kn80 pseudomonas							
139	7	2.9	416	12	Q9JUC4	Q9jqc4 influenza a							
140	7	2.9	419	13	Q92043	Q92043 crocalus at							
141	7	2.9	424	5	Q9GSD3	Q9gsd3 plasmodium							
142	7	2.9	424	5	Q8T9R7	Q8t9r7 plasmodium							
143	7	2.9	424	5	Q9GSD7	Q9gsd7 plasmodium							
144	7	2.9	424	5	Q9NH61	Q9nh61 plasmodium							
145	7	2.9	424	5	Q9NH61	Q9nh61 plasmodium							
146	7	2.9	424	5	Q9NH61	Q9nh61 plasmodium							
147	7	2.9	424	5	Q8GM70	Q8gm70 plasmodium							
148	7	2.9	424	5	Q8GM68	Q8gm68 plasmodium							
149	7	2.9	425	10	Q9ST79	Q9st79 cryza sativ							
150	7	2.9	429	12	Q9Q0L4	Q9q0l4 influenza a							
151	7	2.9	438	12	Q9Q0L3	Q9q0l3 influenza a							
152	7	2.9	438	16	Q99V88	Q99vd8 staphylococ							
153	7	2.9	438	16	Q8CT88	Q8ct88 staphylococ							
154	7	2.9	442	10	Q9SGX5	Q9sgx5 arabidopsis							
155	7	2.9	449	17	Q9PT71	Q9pt71 methanosarc							
156	7	2.9	450	16	Q7TWY4	Q7twy4 mycobacteri							
157	7	2.9	451	12	Q98173	Q98173 molluscum c							
158	7	2.9	451	12	Q98173	Q98173 molluscum c							
159	7	2.9	459	16	Q8XI91	Q8xi91 clostridium							
160	7	2.9	466	11	Q80WC5	Q80wc5 mus musculu							
161	7	2.9	467	10	Q84JS7	Q84js7 arabidopsis							
162	7	2.9	467	12	Q9E312	Q9e312 influenza a							
						Q9e314 influenza a							

## ALIGNMENTS

## RESULT 1

Q9NYC4	PRELIMINARY; PRT; 245 AA.
ID	Q9NYC4
AC	Q9NYC4; (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Tumstatin (Fragment).
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,
RA	Erickson M.D., Hoffer H., Xiao Y., Stillman I.E., Kalluri R.;
RT	"Distinct anti-tumor properties of a type IV collagen domain derived
RT	from basement membrane.";
RL	J. Biol. Chem. 0:0-0(2000).
DR	EMBL; AF258351; AAF7632.1; -
DR	GO; GO:0005581; C:collagen; IEA.
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR	GO; GO:0003676; F:nucleic acid binding; IEA.
DR	InterPro; IPR001442; ProcollagN4 C.
DR	InterPro; IPR000504; RNA_rec_mot.
DR	Pfam; PF01413; C4; 2.
DR	ProDom; PD003923; ProcollagN4; 1.
DR	SMART; SM00111; C4; 2.
DR	PROSITE; PS00030; RRM_RNP_1; 1.
FT	NON_TER 1

SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;  
Query Match 100.0%; Score 244; DB 4; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.6e-242;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGFTVTHSQTTPSCPTVPLYSQFSLFVQGNQRAHQD 60  
Db 2 GLKGRGDSGSPATWTRGFTVTHSQTTPSCPTVPLYSQFSLFVQGNQRAHQD 61  
QY 61 LGTIGSCLQRTTWPFLFCNVNDYCNFASRNDYSYWLSTPALPMNAPITGRALPEYIS 120  
Db 62 LGTIGSCLQRTTWPFLFCNVNDYCNFASRNDYSYWLSTPALPMNAPITGRALPEYIS 121  
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEGTQALASPGSCLE 180  
Db 122 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEGTQALASPGSCLE 181  
QY 181 EFRASPFLECHGRGTCNYNSYSFSLASLNPFRKPIPTSVKAGELEKIIISRCQVCM 240  
Db 182 EFRASPFLECHGRGTCNYNSYSFSLASLNPFRKPIPTSVKAGELEKIIISRCQVCM 241  
QY 241 KKRH 244  
Db 242 KKRH 245  
RESULT 2  
Q28512 PRELIMINARY; PRT; 212 AA.  
AC Q28512;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciinae;  
OC Cercopitheciinae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different mammals.";  
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; L47280; AAA91861.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON-TER 1  
FT NON-TER 212  
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;  
Query Match 48.0%; Score 117; DB 6; Length 212;  
Best Local Similarity 100.0%; Pred. No. 7.1e-112;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 128 PAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEGTQALASPGSCLEFRASGPF 187  
Db 96 PAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEGTQALASPGSCLEFRASGPF 155  
QY 188 LECHGRGTCNYNSYSFSLASLNPFRKPIPTSVKAGELEKIIISRCQVCMKKRH 244

Db 156 LECHGRGTCNYNSYSFSLASLNPFRKPIPTSVKAGELEKIIISRCQVCMKKRH 212  
RESULT 3  
Q28682 PRELIMINARY; PRT; 203 AA.  
AC Q28682;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Alpha-3 type IV collagen (Fragment).  
GN COL4A3.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Cricetidae;  
OX NCBI\_TaxID=9986;  
RN [1]  
RP TISSUE=Kidney cortex;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbamma I.,  
RA Mason P.J., Pusey C.D.;  
RT "Properties and sequences of the Goodpasture antigen of different mammals.";  
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; L47283; AAA91893.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Collagen.  
FT NON-TER 1  
FT NON-TER 203  
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;  
Query Match 22.1%; Score 54; DB 6; Length 203;  
Best Local Similarity 100.0%; Pred. No. 4.7e-47;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 114 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEG 167  
Db 82 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSPFIMFTSAGSEG 135  
RESULT 4  
Q8CCD6 PRELIMINARY; PRT; 112 AA.  
AC Q8CCD6;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Procollagen (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP TISSUE=Lung;  
RC STRAIN=C57BL/6J; PubMed=12466851;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK033387; BAC28260.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

```
DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
SQ SEQUENCE 112 AA; 12438 MW; PFOFDD3C95ATBF31 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 112;
Best Local Similarity 100.0%; Pred. No. 7.8e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 227
DB 57 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 95

RESULT 5
O63122 PRELIMINARY; PRT; 230 AA.
AC Q63122;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RX MEDLINE=98210005; PubMed=9550634;
RA Ryan J.J., Katbama I., Mason P.J., Pusey C.D., Turner A.N.;
RT "Sequence analysis of the 'Goopasture antigen' of mammals.";
RL Nephrol. Dial. Transplant. 13:602-607(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47281; AAB72238.2; -.
DR GO:0005581; F:extracellular matrix structural constituent; IEA.
DR GO:0005201; F:nucleic acid binding; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Collagen.
KW NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 227
DB 175 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 213

RESULT 6
O61435 PRELIMINARY; PRT; 246 AA.
ID O61435
AC Q61435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney cortex;
RA Turner N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47281; AAB72238.2; -.
DR GO:0005581; F:extracellular matrix structural constituent; IEA.
DR GO:0005201; F:nucleic acid binding; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Collagen.
KW NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 227
DB 175 ECHGRGTCNYNSYSFWLASLNPERMFKPIPTVKAG 213

RESULT 7
O9QZS0 PRELIMINARY; PRT; 1669 AA.
ID O9QZS0
AC Q9QZS0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha 3 collagen IV.
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169387; AAD50449.1; -.
DR PIR; I48302; I48302.
DR MGD; MGI:104688; Col4a3.
DR GO:0005604; C:basement membrane; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
```



```
DR Pfam: PF01413; C4; 2
DR Pfam: PF01391; Collagen; 21.
DR ProDom: PD000007; Cig_helix; 6.
DR ProDom: PD003923; ProcollagenC4; 1.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 16.0%; Score 39; DB 11; Length 1669;
Best Local Similarity 100.0%; Pred. No. 7.9e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYFSLASLNPERMFKPIPSIVKAG 227
DB 1614 ECHGRGTCNYNSYFSLASLNPERMFKPIPSIVKAG 1652

RESULT 8
Q61430
ID Q61430 PRELIMINARY; PRT; 161 AA.
AC Q61430;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha3 chain (Fragment).
GN COL4A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Oberbaumer I.;
RT "Cloning of the Ncl domains fo the minor collagen IV chains of mouse
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and
RT alpha5 (IV) in differentiated teratocarcinoma cells.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DDJ databases.
DR EMBL; X82205; CAA57689.1; -.
DR PIR; S49488; S49488.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001442; Procollagn4_C.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF01413; C4; 2.
DR SMART: SM00111; C4; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT NON_TER 161
FT NON_TER 161
SQ SEQUENCE 161 AA; 17925 MW; 1F59DF6CFB8236C5 CRC64;

Query Match 15.6%; Score 38; DB 11; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYNSYFSLASLNPERMFKPIPSIVKAG 226
DB 124 ECHGRGTCNYNSYFSLASLNPERMFKPIPSIVKAG 161

RESULT 9
Q9UDK9
ID Q9UDK9 PRELIMINARY; PRT; 52 AA.
AC Q9UDK9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Penades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieslander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms.";
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; S69113; S69113.
FT NON_TER 1
FT NON_CONS 45
FT NON_TER 52
FT NON_TER 52
SQ SEQUENCE 52 AA; 5442 MW; 046AB41B149DDAE3 CRC64;

Query Match 14.8%; Score 36; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.9e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLXGKRGDSGPATWTRGTFVTRHSQTTAIPSCPE 36
DB 10 GLXGKRGDSGPATWTRGTFVTRHSQTTAIPSCPE 45

RESULT 10
Q9UDLO
ID Q9UDLO PRELIMINARY; PRT; 56 AA.
AC Q9UDLO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Penades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieslander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms.";
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; A49736; A49736.
FT NON_TER 1
FT NON_CONS 45
FT NON_TER 56
FT NON_TER 56
SQ SEQUENCE 56 AA; 5813 MW; 6A6A2E22819F473B CRC64;

Query Match 14.8%; Score 36; DB 4; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.2e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLXGKRGDSGPATWTRGTFVTRHSQTTAIPSCPE 36
DB 10 GLXGKRGDSGPATWTRGTFVTRHSQTTAIPSCPE 45

RESULT 11
Q9UDL1
ID Q9UDL1 PRELIMINARY; PRT; 97 AA.
AC Q9UDL1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GOODPASTURE antigen (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95278230; PubMed=7758473;
RA Penades J.R., Bernal D., Revert F., Johansson C., Fresquet V.J.,
RA Cervera J., Wieselander J., Quinones S., Saus J.;
RT "Characterization and expression of multiple alternatively spliced
RT transcripts of the Goodpasture antigen gene region. Goodpasture
RT antibodies recognize recombinant proteins representing the autoantigen
RT and one of its alternative forms."
RL Eur. J. Biochem. 229:754-760(1995).
DR PIR; B49736; B49736.
FT NON_TER 1
FT NON_CONS 45
FT NON_TER 97
SQ SEQUENCE 97 AA; 10559 MW; B3CC1127F24A27E3 CRC64;

Query Match 14.8%; Score 36; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.4e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPE 36
DB 10 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPE 45

RESULT 12
Q28567 PRELIMINARY; PRT; 212 AA.
AC Q28567;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47282; AAA91904.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR PIR; PF01413; C4; 2.
DR Pfam; PF000504; RNA_rec_mot.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 14.3%; Score 35; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.8e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 GWISLWKGFSTFMTSAGSEGTGQALAGPGSCLEE 181
DB 115 GWISLWKGFSTFMTSAGSEGTGQALAGPGSCLEE 149
```

```
RESULT 13
Q28273 PRELIMINARY; PRT; 210 AA.
AC Q28273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 3 chain (Fragment).
GN COL4A3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation."
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50935; AAC48585.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 210
SQ SEQUENCE 210 AA; 23025 MW; 31119B4CA823633D CRC64;

Query Match 10.2%; Score 25; DB 6; Length 210;
Best Local Similarity 100.0%; Pred. No. 3.4e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 AHGQDLGTGSCLOQRTTTPFLFCN 80
DB 34 AHGQDLGTGSCLOQRTTTPFLFCN 58

RESULT 14
Q29032 PRELIMINARY; PRT; 203 AA.
AC Q29032;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-3 type IV collagen (Fragment).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney cortex;
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RT "Properties and sequences of the Goodpasture antigen of different
RT mammals."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47284; AAA91862.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01413; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 23025 MW; 31119B4CA823633D CRC64;
```

DR ProDom: PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RNP\_RNP\_1; 1.  
KW Collagen.  
FT NON\_TER 1 1  
FT NON\_TER 203 203  
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 9.8%; Score 24; DB 6; Length 203;  
Best Local Similarity 100.0%; Pred. No. 3.5e-16;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 HGQDLGTGSCLORETTMPFLFCN 80  
|||||  
DB 25 HGQDLGTGSCLORETTMPFLFCN 48  
|||||

## RESULT 15

Q9UDJ9 PRELIMINARY; PRT; 29 AA.  
AC Q9UDJ9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE GOODPASTURE antigen (Fragments).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced".  
RL J. Biol. Chem. 268:12090-12094 (1993).  
FT NON\_TER 1 1  
FT NON\_CONS 18 19  
FT NON\_TER 29 29  
SQ SEQUENCE 29 AA; 3102 MW; 2B7047AAB1580036 CRC64;

Query Match 7.4%; Score 18; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTR 18  
|||||  
DB 1 GLKGRGDSGSPATWTR 18  
|||||

## RESULT 16

P70165 PRELIMINARY; PRT; 179 AA.  
ID P70165;  
AC P70165;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha5 chain (fragment).  
GN COL4A5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Oberbauer I.;  
RT "Cloning of the NC1 domains of the minor collagen IV chains of mouse  
RT via PCR (RACE) reveals the presence of the mRNAs for alpha3 (IV) and  
RT alpha5 (IV) in differentiated teratocarcinoma cells.";  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X82218; CAA57698.1; .  
GO; GO:0005581; C:collagen; IEA.

DR GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
FT NON\_TER 1 1  
FT NON\_TER 179 179  
SQ SEQUENCE 179 AA; 19859 MW; 20A188F3687F582F CRC64;

Query Match 7.0%; Score 17; DB 11; Length 179;  
Best Local Similarity 100.0%; Pred. No. 5.1e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
|||||  
DB 47 VCNFASRNDYSYWLSTP 63  
|||||

## RESULT 17

Q28271 PRELIMINARY; PRT; 225 AA.  
ID Q28271;  
AC Q28271;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha 1 chain (Fragment).  
GN COL4A1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96278820; PubMed=8662866;  
RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;  
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains  
RT of collagen type IV. Evidence from a canine model of X-linked  
RT nephritis with a COL4A5 gene mutation.";  
RL J. Biol. Chem. 271:13821-13828 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA STRAIN=Samoyed;  
RX MEDLINE=96278820; PubMed=8662866;  
RA Thorner P.S.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U50933; AAC48583.2; .  
GO; GO:0005581; C:collagen; IEA.  
DR GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; ProcollagnC4; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER 1 1  
FT NON\_TER 225 225  
SQ SEQUENCE 225 AA; 24585 MW; 2C20455890416E47 CRC64;

Query Match 7.0%; Score 17; DB 6; Length 225;  
Best Local Similarity 100.0%; Pred. No. 6.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
|||||  
DB 76 VCNFASRNDYSYWLSTP 92  
|||||

## RESULT 18

Q99LQ8 PRELIMINARY; PRT; 226 AA.  
ID Q99LQ8;  
AC Q99LQ8;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

```

DE Hypothetical protein (Fragment).
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002269; AA02269.1; -.
DR MGD; MGI:88454; Col4a1.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 226 AA; 25042 MW; 4F7F0D5371181C21 CRC64;

Query Match 7.0%; Score 17; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 67 VCNFASRNDYSYWLSTP 83
|||||

RESULT 19
Q9NF88 PRELIMINARY; PRT; 229 AA.
AC Q9NF88;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA He A.B.;
RL "Cloning and Expression of Arresten in Escherichia coli and Pachia
pastoris.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF536207; AA097359.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 229 AA; 25391 MW; 09B21F5A351789E CRC64;

Query Match 7.0%; Score 17; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 70 VCNFASRNDYSYWLSTP 86
|||||

RESULT 20
Q9NYC5 PRELIMINARY; PRT; 229 AA.
AC Q9NYC5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arresten (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,
Takanashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K.,
Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D.,
Weichselbaum R.R., Sukhatme V.P., Kalluri R.;
RT "Anti-angiogenic cues from vascular basement membrane collagen.";
RN [2]
RP SEQUENCE FROM N.A.
RA Fu J., Bai X., Wang W., Ruan C.;
RT "Arresten, a collagen-derived inhibitor of angiogenesis.";
RL Chung Hua Hsueh Yeh Hsueh Tsa Chih 22:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Peng X., Yin B., Yuan J., Qiang B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Zheng Q.C., Song Z.F., Zheng Y.W., Li Y.Q., Shu X.;
RT "Molecular cloning and sequencing of human arresten gene.";
RL Zhonghua Shi Yan Wai Ke Za Zhi 19:46-47(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Song Z.F., Zheng Q.C.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258349; AA072630.1; -.
DR EMBL; AF363672; AA033382.1; -.
DR EMBL; AF400431; AA092480.1; -.
DR EMBL; AV285780; AA043112.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CDC100A5C1D5 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 70 VCNFASRNDYSYWLSTP 86
|||||

RESULT 21
Q61436 PRELIMINARY; PRT; 253 AA.
AC Q61436;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen IV alpha 5 chain (Fragment).
GN COL4A5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Muscle;
RC MEDLINE=95050957; PubMed=7962065;
RA Miner J.H., Sane J.R.;
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal

```

DR SMART; SM00111; C4; 2  
 SQ SEQUENCE 585 AA; 58283 MW; 26774FE364F7FD8D CRC64;

Query Match 7.0%; Score 17; DB 11; Length 585;  
 Best Local Similarity 100.0%; Pred.No.1.4e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
 |||||||  
 DB 426 VCNFASRNDYSYWLSTP 442

RESULT 23  
 Q8BNS7 PRELIMINARY; PRT; 799 AA.

ID Q8BNS7  
 AC Q8BNS7; 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Procollagen (Fragment).  
 GN COL4A5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cortex;  
 RX MEDLINE=22354883; PubMed=12456851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 DR ENBL; AK080682; BAC37980.1; -  
 DR MGD; MGI:88456; Col4a5.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagn4\_C.  
 DR PFAM; PF01413; C4; 2.  
 DR PFAM; PF01391; Collagen; 9.  
 DR ProDom; PD000007; C1g\_helix; 1.  
 DR ProDom; PD003923; ProcollagnC4; 1.  
 DR SMART; SM00111; C4; 2.  
 FT NON TER 1  
 SQ SEQUENCE 799 AA; 77889 MW; C517CF4CF15706DC CRC64;

Query Match 7.0%; Score 17; DB 11; Length 799;  
 Best Local Similarity 100.0%; Pred.No.1.8e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
 |||||||  
 DB 640 VCNFASRNDYSYWLSTP 656

RESULT 24  
 Q9NJB7 PRELIMINARY; PRT; 886 AA.

ID Q9NJB7  
 AC Q9NJB7  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE D242423.1 (Collagen, type IV, alpha 5 (Alport syndrome))  
 DE (Fragment).  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

```
EA Copley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 10.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9A86569 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 886;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 727 VCNFASRNDYSYWLSTP 743
|||||

RESULT 25
Q919X3 PRELIMINARY; PRT; 979 AA.
AC Q919X3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Collagen IV al chain (Fragment).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
RT basal lamina.";
RL Dev. Biol. 0:0-0(2000).
DR EMBL; AF239838; AAF4681.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 12.
DR ProDom; PD000007; Clg helix; 2.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 7.0%; Score 17; DB 13; Length 979;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 820 VCNFASRNDYSYWLSTP 836
|||||

RESULT 26
Q86X41 PRELIMINARY; PRT; 1075 AA.
ID Q86X41
AC Q86X41;
```

```
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to collagen, type IV, alpha 1 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047305; AAH47305.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1075 AA; 103426 MW; 4802654BD552503D CRC64;

Query Match 7.0%; Score 17; DB 4; Length 1075;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100
DB 916 VCNFASRNDYSYWLSTP 932
|||||

RESULT 27
Q9H4R9 PRELIMINARY; PRT; 1621 AA.
ID Q9H4R9;
AC Q9H4R9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE BA472K17.2 (Collagen type IV alpha 1 (fragment)).
DE COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390755; CAC13153.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; Clg helix; 5.
DR ProDom; PD003923; ProcollagnC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1621 AA; 155705 MW; 73F6F901CDD0EBA2 CRC64;

Query Match 7.0%; Score 17; DB 4; Length 1621;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      84 VCNFASRNDYSYWLSTP 100
Db      1462 VCNFASRNDYSYWLSTP 1478

RESULT 28
QBHYC1
ID      Q8HYC1      PRELIMINARY;      PRT; 1684 AA.
AC      Q8HYC1;
DT      01-WAR-2003 (TREMBlrel. 23, Created)
DT      01-WAR-2003 (TREMBlrel. 23, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Type IV collagen alpha 5 chain (Fragment).
GN      COL4A5.
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Harvey S.J., Zheng K., Jefferson B., Sado Y., Naito I., Ninomiya Y.,
RA      Jacobs R., Thorne P.S.;
RT      Recombinant alpha5(IV) collagen: In vivo adenoviral-mediated gene
RT      transfer to smooth muscle restores expression of the alpha6(IV)
RT      collagen chain in a canine model of Alport syndrome."
RL      Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY078501; AAL83712.1; -.
DR      GO; GO:0005581; C:collagen; IEA.
DR      GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR      InterPro; IPR008161; C:lg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 26.
DR      ProDom; PD000007; Clg_helix; 3.
DR      ProDom; PD003923; ProcollagnC4; 1.
DR      SMART; SM00111; C4; 2.
KW      Collagen.
FT      NON_TER      1684
SQ      SEQUENCE      1684 AA; 161408 MW; 02D631B545F285D CRC64;

Query Match      7.0%; Score 17; DB 6; Length 1684;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 29
QB66Z2
ID      Q866Z2      PRELIMINARY;      PRT; 1688 AA.
AC      Q866Z2;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Type IV collagen alpha 5.
GN      COL4A5.
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Cox M.L., Lees G.E., Kashtan C.E., Murphy K.E.;
RT      "Genetic Cause of X-linked Alport Syndrome in a Family of Domestic
RT      Dogs."
RL      Mamm. Genome 0:0-0(2003).
DR      EMBL; AF470624; AAO33458.1; -.
DR      GO; GO:0005581; C:collagen; IEA.
DR      GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
```

```
DR      InterPro; IPR008161; Clg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 21.
DR      ProDom; PD000007; Clg_helix; 2.
DR      ProDom; PD003923; ProcollagnC4; 2.
DR      SMART; SM00111; C4; 2.
KW      Collagen.
SQ      SEQUENCE      1688 AA; 161725 MW; 7121BE329931CDBEC CRC64;

Query Match      7.0%; Score 17; DB 6; Length 1688;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 30
Q9ESQ2
ID      Q9ESQ2      PRELIMINARY;      PRT; 1691 AA.
AC      Q9ESQ2;
DT      01-MAR-2001 (TREMBlrel. 16, Created)
DT      01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Type IV collagen alpha 5 chain.
GN      COL4A5.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20536494; PubMed=10955041;
RA      Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
RA      Kishiro Y., Sado Y., Yoshio H., Ninomiya Y.;
RT      "Differential Expression of Mouse alpha5(IV) and alpha6(IV) Collagen Genes in
RT      Epithelial Basement Membranes."
RL      J. Biochem. 128:427-434(2000).
DR      EMBL; AB041350; BAB13673.1; -.
DR      MGD; MGI:88456; Col4a5.
DR      GO; GO:0005604; C:basement membrane; IEA.
DR      InterPro; IPR008161; Clg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001442; Procollagn4_C.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 24.
DR      ProDom; PD000007; Clg_helix; 3.
DR      ProDom; PD003923; ProcollagnC4; 1.
DR      SMART; SM00111; C4; 2.
KW      Collagen.
SQ      SEQUENCE      1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match      7.0%; Score 17; DB 11; Length 1691;
Best Local Similarity 100.0%; Pred.No. 3.5e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 VCNFASRNDYSYWLSTP 100
Db      1532 VCNFASRNDYSYWLSTP 1548

RESULT 31
Q07265
ID      Q07265      PRELIMINARY;      PRT; 1752 AA.
AC      Q07265;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      3 alpha procollagen.
GN      COL3ALPHA.
```



OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=8444899;  
RX MEDLINE=93186842; Di Liberto M., Ramirez F.;  
RA Exposito J.-Y., D'Alessio M., Di Liberto M., Ramirez F.;  
RT "Complete primary structure of a sea-urchin type IV collagen and  
analysis of the 5' end of its gene."  
RL J. Biol. Chem. 268:5249-5254(1993).  
DR EMBL: L02917; AAA30039.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 25.  
DR ProDom; PD000007; C1g\_helix; 9.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1752 AA; 170210 MW; 1AESAAR21569346D CRC64;  
  
Query Match 4.9%; Score 12; DB 5; Length 1752;  
Best Local Similarity 100.0%; Pred. No. 0.005;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 88 ASRNDYSYWLST 99  
Db 1597 ASRNDYSYWLST 1608  
|||||  
  
RESULT 32  
Q9GV24 PRELIMINARY; PRT; 713 AA.  
ID Q9GV24  
AC Q9GV24  
DT 01-WAR-2001 (TrEMBLrel. 16, Created)  
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha-2 (fragment).  
OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
OC Sarcophagidae; Sarcophaga.  
OX NCBI\_TaxID=7386;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20356508; PubMed=10965055;  
RA Fujii-Taira I., Tanaka Y., Homma K.J., Natori S.;  
RT "Hydrolysis and synthesis of substrate proteins for cathepsin L in the  
brain basement membranes of Sarcophaga during metamorphosis."  
RL J. Biochem. 128:539-542(2000).  
DR EMBL; AB041728; BAB1607.1;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 6.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
FT NON TER 1  
SQ SEQUENCE 713 AA; 74704 MW; 122A361259E40DCB CRC64;  
  
Query Match 4.5%; Score 11; DB 5; Length 713;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 174 SPGSCLEEFRA 184  
|||||

Db 573 SPGSCLEEFRA 583  
  
RESULT 33  
O18407 PRELIMINARY; PRT; 1761 AA.  
ID O18407  
AC O18407  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen type IV alpha 2.  
GN VKG OR DMCOLA2 OR CG16858.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yasothornsikul S., Davis W.J., Cramer G., Kimbrell D.A.,  
RA Dearolf C.R.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55431; AA864082.1; -.  
DR FIR; T13990; T13990.  
DR FlyBase; FBgn0016075; vkg.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 25.  
DR ProDom; PD000007; C1g\_helix; 12.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1761 AA; 175955 MW; FCB23AFF19121DC6 CRC64;  
  
Query Match 4.5%; Score 11; DB 5; Length 1761;  
Best Local Similarity 100.0%; Pred. No. 0.054;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 174 SPGSCLEEFRA 184  
Db 1668 SPGSCLEEFRA 1678  
|||||  
  
RESULT 34  
Q9VMV5 PRELIMINARY; PRT; 1940 AA.  
ID Q9VMV5  
AC Q9VMV5  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE VKG protein.  
GN VKG OR CG16858.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Berkely.  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dalkic C., Davenport L.B., Davies P.,  
RA De Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Rulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moehrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003608; AAF52203.1; -.  
DR Flybase; Fgn0016075; vkg.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 7.  
DR ProDom; PD000007; Clg\_helix; 12.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1940 AA; 193777 MW; 9B507382EF9C17B5 CRC64;  
Query Match 4.5%; Score 11; DB 5; Length 1940;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 174 SPGSCLEPRA 184  
Db 1670 SPGSCLEPRA 1680  
RESULT 35  
ID Q14052 PRELIMINARY; PRT; 573 AA.  
AC Q14052;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha-2 type IV collagen (Fragment).  
GN COL4A2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88085168; PubMed=3692475;  
RA Killen P.D., Francomano C.A., Yamada Y., Modi W.S., O'Brien S.J.;  
RT "Partial structure of the human alpha 2(IV) collagen chain and  
RT chromosomal localization of the gene (COL4A2).";  
RL Hum. Genet. 77:318-324(1987).

DR EMBL; M24756; AAA52043.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 7.  
DR ProDom; PD000007; Clg\_helix; 3.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON TER  
SQ SEQUENCE 673 AA; 67174 MW; D2F3C9B3111A3105 CRC64;  
Query Match 4.1%; Score 10; DB 4; Length 673;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 128 PAIAIAVHSQ 137  
Db 557 PAIAIAVHSQ 566  
RESULT 36  
ID Q8T7S4 PRELIMINARY; PRT; 1024 AA.  
AC Q8T7S4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen IV alpha 1 chain (Fragment).  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gare D.C., Billingsley P.F.;  
RT "Mquito collagen IV: conservation of the NC1 domain between alpha  
RT chains.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF313909; AAL99382.1; -.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 13.  
DR ProDom; PD000007; Clg\_helix; 2.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON TER  
SQ SEQUENCE 1024 AA; 101502 MW; 6BC10D7219C44D68 CRC64;  
Query Match 4.1%; Score 10; DB 5; Length 1024;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 117 PYISRCTVCE 126  
Db 898 PYISRCTVCE 907  
RESULT 37  
ID Q26640 PRELIMINARY; PRT; 1747 AA.  
AC Q26640;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Alpha2(IV)-like collagen.  
GN COLP4ALPHA.  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoida; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94230414; PubMed=8175744;  
RA Exposito J.Y., Suzuki H., Georjon C., Garrone R., Solursh M.,  
RA Ramirez F.;  
RT "Identification of a cell lineage-specific gene coding for a sea  
RT urchin alpha.2(IV)-like collagen chain.";  
RL J. Biol. Chem. 269:13167-13171(1994).  
DR EMBL; X76730; CAA54146.1; -.  
DR PIR; A54121; A54121.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR ProDom; PD000007; C1g\_helix; 6.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Collagen.  
KW Collagen.  
SQ SEQUENCE 1747 AA; 173312 MW; EE722E878394B9B6 CRC64;  
Query Match 4.1%; Score 10; DB 5; Length 1747;  
Best Local Similarity 100.0%; Pred. No. 0.58;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 117 PYISRCTVCE 126  
Db 1622 PYISRCTVCE 1631  
RESULT 38  
QBP19 PRELIMINARY; PRT; 59 AA.  
ID Q8BP19;  
AC Q8BP19;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Procollagen (Fragment).  
GN COL4A2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Body;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK075619; BAC35863.1; -.  
DR MGD; MGI:88455; Col4a2.  
DR GO; GO:0005604; C:basement membrane; IEA.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 1.  
DR SMART; SM00111; C4; 1.  
DR NON\_TER 1  
FT NON\_TER  
SQ SEQUENCE 59 AA; 6697 MW; 61F00BA79B3B4566 CRC64;  
Query Match 3.7%; Score 9; DB 11; Length 59;  
Best Local Similarity 100.0%; Pred. No. 0.34;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ISRCQVCMK 241  
Db 49 ISRCQVCMK 57  
RESULT 39  
Q64457 PRELIMINARY; PRT; 312 AA.  
ID Q64457;  
AC Q64457;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Collagen IV alpha 4 chain (Fragment).  
GN COL4A4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Balb/C; TISSUE=Kidney;  
RX MEDLINE=95050957; PubMed=7962065;  
RA Miner J.H., Sanes J.R.;  
RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal  
RT laminae: Sequence, distribution, association with laminins, and  
RT developmental switches.";  
RL J. Cell Biol. 127:879-891(1994).  
DR EMBL; Z35167; CAA84530.1; -.  
DR PIR; I48303; I48303.  
DR MGD; MGI:104687; Col4a4.  
DR GO; GO:0005604; C:basement membrane; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 1.  
DR ProDom; PD000007; C1g\_helix; 1.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Collagen.  
KW NON\_TER  
FT NON\_TER  
SQ SEQUENCE 312 AA; 33132 MW; EB017D02868C681E CRC64;  
Query Match 3.7%; Score 9; DB 11; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 233 ISRCQVCMK 241  
Db 302 ISRCQVCMK 310  
RESULT 40  
Q91V13 PRELIMINARY; PRT; 358 AA.  
ID Q91V13;  
AC Q91V13;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN COL4A2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Breast tumor;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC013560; AAH13560.1; -.  
DR MGD; MGI:88455; Col4a2.

DR GO: 0005604; C:basement membrane; IDA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 2.  
DR ProDom; PD000007; Clg\_helix; 1.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2. Collagen.  
KW Hypothetical protein; Collagen.  
SQ SEQUENCE 358 AA; 37432 MW; 6DA02BDF3062D9E CRC64;  
  
Query Match 3.7%; Score 9; DB 11; Length 358;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 233 ISRCQVCMK 241  
DB 348 ISRCQVCMK 356  
|||||  
RESULT 41  
Q9QZR9 PRELIMINARY; PRT; 1682 AA.  
AC Q9QZR9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Alpha 4 collagen IV.  
GN COL4A4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]\_TaxID=10090;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=20005934; PubMed=10534397;  
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,  
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;  
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a  
RT mouse model of alport syndrome.";  
RL Genomics 61:113-124(1999).  
DR EMBL; AF169388; AD50450.1; -.  
DR MGD; MGI:104687; Col4a4.  
DR GO: 0005604; C:basement membrane; IDA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 22.  
DR ProDom; PD000007; Clg\_helix; 4.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76B904 CRC64;  
  
Query Match 3.7%; Score 9; DB 11; Length 1682;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 233 ISRCQVCMK 241  
DB 1672 ISRCQVCMK 1680  
|||||  
RESULT 42  
Q48996 PRELIMINARY; PRT; 74 AA.  
AC Q48996;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE 1-phosphofructinase (Fragment).  
OS Mycoplasma capricolum.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 27343 (K1d);  
RX MEDLINE=96059641; PubMed=7476192;  
RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,  
RA Gilbert W., Gillevet P.M.;  
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals  
RT its physiology.";  
RL Mol. Microbiol. 16:955-967(1995).  
DR EMBL; Z33072; CAA83740.1; -.  
DR PUR; S77782; S77782.  
FT NON\_TER 1  
FT NON\_TER 74  
SQ SEQUENCE 74 AA; 8281 MW; 3AD85E191264D90F CRC64;  
  
Query Match 3.3%; Score 8; DB 2; Length 74;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 197 NYYSNSYS 204  
DB 26 NYYSNSYS 33  
|||||  
RESULT 43  
Q8T2W5 PRELIMINARY; PRT; 179 AA.  
AC Q8T2W5;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Tccl114-1.2.  
GN Tccl114-1.2.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL Brenner;  
RA Anderson B., Bontempi E.J.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC114396; AAL86598.1; -.  
DR GO: 0016020; C:membrane; IEA.  
DR GO: 0005215; F:transporter activity; IEA.  
DR GO: 0006810; P:transport; IEA.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR005829; Sug\_transporter.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
SQ SEQUENCE 179 AA; 19990 MW; AE40862E2018E357 CRC64;  
  
Query Match 3.3%; Score 8; DB 5; Length 179;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 45 FSFLFVQG 52  
DB 14 FSFLFVQG 21  
|||||  
RESULT 44  
Q81T03 PRELIMINARY; PRT; 304 AA.  
AC Q81T03;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Superoxide dismutase.  
GN BA1489.

OS Bacillus anthracis (strain Ames).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=198094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608414; PubMed=12721629;  
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
RA Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria."  
RL Nature 423:81-86(2003).  
DR EMBL; AF017028; AAP25428.1; --  
DR TIGR; BA1489; --  
DR GO; GO:0046872; F-metal ion binding; IEA.  
DR GO; GO:0004784; F-superoxide dismutase activity; IEA.  
DR GO; GO:0005801; P-superoxide metabolism; IEA.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR Pfam; PF02777; sodfe C; 1.  
DR PRINTS; PRO1703; MNSODISMUTASE.  
DR ProDom; PD000475; SODismutase; 1.  
DR PROSITE; PS00088; SOD\_MN; 1.  
KW Complete proteome.  
SQ SEQUENCE 304 AA; 36161 MW; 4CBA6BA31B46D1FD CRC64;  
  
Query Match 3.3%; Score 8; DB 16; Length 304;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 114 ALEPYISR 121  
DB 114 ALEPYISR 121  
|||||  
  
RESULT 45  
Q8BX78 PRELIMINARY; PRT; 337 AA.  
AC Q8BX78;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to mitochondrial carrier family protein.  
GN B230315F11RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Retina;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK044260; BAC31842.1; --  
DR MGD; MGI:2442486; B230315F11RIK.  
DR GO; GO:0005743; C-mitochondrial inner membrane; IEA.  
DR GO; GO:0005488; F-binding; IEA.  
DR GO; GO:0006810; P-transport; IEA.  
DR InterPro; IPR001993; Mitoch.carrier.  
DR InterPro; IPR002067; Mit.carrier.  
DR Pfam; PF00153; mito\_carr; 3.  
DR PRINTS; PR00926; MITOCARRIER.  
DR PROSITE; PS00215; MITOCH\_CARRIER; 1.  
SQ SEQUENCE 337 AA; 37963 MW; 80554124DD995581 CRC64;  
  
Query Match 3.3%; Score 8; DB 11; Length 337;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 147 GWISLWKG 154  
DB 192 GWISLWKG 199  
|||||  
  
RESULT 47  
Q921S0 PRELIMINARY; PRT; 452 AA.  
AC Q921S0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to hypothetical protein FLJ12810.  
GN DCLEIB OR A145214.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

DR PRINTS; PR00926; MITOCARRIER.  
DR PROSITE; PS00215; MITOCH\_CARRIER; 1.  
SQ SEQUENCE 337 AA; 38023 MW; 8948A7D6DB624861 CRC64;  
  
Query Match 3.3%; Score 8; DB 11; Length 337;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 147 GWISLWKG 154  
DB 192 GWISLWKG 199  
|||||  
  
RESULT 46  
Q8BGP6 PRELIMINARY; PRT; 337 AA.  
AC Q8BGP6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to mitochondrial carrier family protein.  
GN B230315F11RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; and NOD; TISSUE=Brain, Heart, and Thymus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK045863; BAC32513.1; --  
DR EMBL; AK084797; BAC39281.1; --  
DR EMBL; AK088444; BAC40356.1; --  
DR MGD; MGI:2442486; B230315F11RIK.  
DR GO; GO:0005743; C-mitochondrial inner membrane; IEA.  
DR GO; GO:0005488; F-binding; IEA.  
DR GO; GO:0006810; P-transport; IEA.  
DR InterPro; IPR001993; Mitoch.carrier.  
DR InterPro; IPR002067; Mit.carrier.  
DR Pfam; PF00153; mito\_carr; 3.  
DR PRINTS; PR00926; MITOCARRIER.  
DR PROSITE; PS00215; MITOCH\_CARRIER; 1.  
SQ SEQUENCE 337 AA; 37963 MW; 80554124DD995581 CRC64;  
  
Query Match 3.3%; Score 8; DB 11; Length 337;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 147 GWISLWKG 154  
DB 192 GWISLWKG 199  
|||||  
  
RESULT 47  
Q921S0 PRELIMINARY; PRT; 452 AA.  
AC Q921S0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to hypothetical protein FLJ12810.  
GN DCLEIB OR A145214.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; SC01094; AH01094.1; -;  
 DR MGD; MGI:2156057; Dclrelb.  
 SQ SEQUENCE 452 AA; 51011 MW; 454DA78A3C235ABD CRC64;

Query Match 3.3%; Score 8; DB 11; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13  
 Db 370 RGDGSGSPA 377

## RESULT 48

QBN95 PRELIMINARY; PRT; 486 AA.  
 AC QBN95;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE DNA cross-link repair 1B.  
 GN DCLRE1B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Eye;  
 RC MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK084347; BAC39165.1; -;  
 DR MGD; MGI:2156057; Dclrelb.  
 SQ SEQUENCE 486 AA; 54622 MW; E9F3CFAC184A9D3F CRC64;

Query Match 3.3%; Score 8; DB 11; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13  
 Db 404 RGDGSGSPA 411

## RESULT 49

Q8THZ6 PRELIMINARY; PRT; 523 AA.  
 AC Q8THZ6;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein MA4362.  
 GN MA4362.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RC MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA FitzHugh W., Calvo S., Engle R., Smirnov S., Atnoor D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,  
 RA Linton L., McSwan P., McKernan K., Talamas J., Firrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,

RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AB011155; AF007704.1; -;  
 DR GO; GO:0000155; Ftwo-component sensor molecule activity; IEA.  
 DR GO; GO:0000160; Ptwo-component signal transduction system (P. . .); IEA.  
 DR InterPro; IPR003018; GAF.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000700; PAS-assoc C.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF01590; GAF; 1.  
 DR Pfam; PF00785; PAC; 2.  
 DR SMART; SM00065; GAF; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR TIGRFAMs; TIGR00229; sensory\_box; 2.  
 DR PROSITE; PS01113; PAC; 2.  
 DR PROSITE; PS01112; PAS; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 523 AA; 59084 MW; C9FF2C5CC0AF32C CRC64;

Query Match 3.3%; Score 8; DB 17; Length 523;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 AIAIAVHS 136  
 Db 13 AIAIAVHS 20

## RESULT 50

Q8C7W7 PRELIMINARY; PRT; 541 AA.  
 AC Q8C7W7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE DNA cross-link repair 1B.  
 GN DCLRE1B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J;  
 RC MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK049115; BAC33550.1; -;  
 DR MGD; MGI:2156057; Dclrelb.  
 SQ SEQUENCE 541 AA; 61055 MW; 281C6AF82F01B76D CRC64;

Query Match 3.3%; Score 8; DB 11; Length 541;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RGDGSGSPA 13  
 Db 459 RGDGSGSPA 466

## RESULT 51

Q99K97 PRELIMINARY; PRT; 546 AA.  
 ID Q99K97

AC Q99K97; 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN COL4A6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC004800; AAH04800.1; -.  
DR MGD; MGI:2152695; Col4a6.  
DR GO; GO:0005587; C:collagen type IV; IDA.  
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; IDA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 5.  
DR ProDom; PD000007; C1g\_helix; 1.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Hypothetical protein; Collagen.  
FT NON TER 1  
SQ SEQUENCE 546 AA; 55102 MW; 56F8CC69374BBCFE CRC64;  
Query Match 3.3%; Score 8; DB 11; Length 546;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 234 SRCQVCVK 241  
DB 537 SRCQVCVK 544  
RESULT 52  
Q9C250  
ID Q9C250 PRELIMINARY; PRT; 1477 AA.  
AC Q9C250;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Related to multidrug resistance-associated protein.  
GN B18D24.240.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL513466; CAC28822.2; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR001140; ABC TM transp.  
DR InterPro; IPR003439; ABC transporter.  
DR Pfam; PF00664; ABC membrane; 1.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR SMART; SM00382; AAA; 2.

DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
KW ATP-binding; 1477 AA; 161202 MW; 28701BF2084EAB75 CRC64;  
SQ SEQUENCE 1477 AA; 161202 MW; 28701BF2084EAB75 CRC64;  
Query Match 3.3%; Score 8; DB 3; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 158 IMFTSAGS 165  
DB 984 IMFTSAGS 991  
RESULT 53  
Q9ESQ1  
ID Q9ESQ1 PRELIMINARY; PRT; 1691 AA.  
AC Q9ESQ1;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Type IV collagen alpha 6 chain.  
GN COL4A6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=20536494; PubMed=10955041;  
RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,  
RA Kishiro Y., Sado Y., Yoshioka H., Ninomiya Y.;  
RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in  
RT Epithelial Basement Membranes.";  
RL J. Biochem. 128:427-434 (2000).  
DR EMBL; AB041351; BAB13674.1; -.  
DR MGD; MGI:2152695; Col4a6.  
DR GO; GO:0005587; C:collagen type IV; IDA.  
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; IDA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 22.  
DR ProDom; PD000007; C1g\_helix; 7.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1691 AA; 164145 MW; CA7E4031DF04F7A7 CRC64;  
Query Match 3.3%; Score 8; DB 11; Length 1691;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 234 SRCQVCVK 241  
DB 1682 SRCQVCVK 1689  
RESULT 54  
Q9GQB1  
ID Q9GQB1 PRELIMINARY; PRT; 1723 AA.  
AC Q9GQB1;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Type IV collagen alpha 1 chain precursor.  
OS Hydra attenuata (Hydra) (Hydra vulgaris).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;  
OC Hydridae; Hydra.  
OX NCBI\_TaxID=6087;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564332; PubMed=10956657;



RA Fowler S.J., Jose S., Zhang X., Deutzmann R., Sarraz M.P. Jr.,  
RA Boot-Handford R.P., Zhang X., Deutzmann R., Sarraz M.P. Jr.,  
RT "Characterization of hydra type IV collagen: Type IV collagen is  
RT essential for head regeneration and its expression is up-regulated  
RT upon exposure to glucose";  
RL J. Biol. Chem. 275:39589-39599 (2000).  
DR EMBL; AF282902; AAG40729.1; -  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; C1g\_helix; 6.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen; Signal.  
FT SIGNAL 1 24 POTENTIAL.  
SQ SEQUENCE 1723 AA; 168996 MW; 92496D62FD162F01 CRC64;  
  
Query Match 3.3%; Score 8; DB 5; Length 1723;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 189 ECHGRGTC 196  
DB 1669 ECHGRGTC 1676  
|||||  
RESULT 55  
Q9VMV4 PRELIMINARY; PRT; 1779 AA.  
ID Q9VMV4  
AC Q9VMV4  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE CG25C protein.  
GN CG25C OR C04145.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
CX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA Sutton G.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195 (2000).  
DR EMBL; AF003608; AAF52204.1; -  
DR FlyBase; FBgn000299; Cg25C.  
DR GO; GO:0005587; C:collagen type IV; NAS.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; C1g\_helix; 9.  
DR ProDom; PD003923; ProcollagnC4; 1.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1779 AA; 174300 MW; 6770F18AB40A313B CRC64;  
  
Query Match 3.3%; Score 8; DB 5; Length 1779;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 234 SRCQVCWK 241  
DB 1769 SRCQVCWK 1776  
|||||  
RESULT 56  
Q9R4U1 PRELIMINARY; PRT; 49 AA.  
ID Q9R4U1  
AC Q9R4U1  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Iron superoxide dismutase, FE-SOD (EC 1.15.1.1) (Superoxide dismutase  
DE [Mn/Fe] (Fragment)).  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
CX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=95094938; PubMed=8001685;  
RA Pagani S., Colnaghi R., Palagi A., Negri A.;  
RT "Purification and characterization of an iron superoxide dismutase  
RT from the nitrogen-fixing Azotobacter vinelandii";  
RL FEBS Lett. 357:79-82 (1995).  
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELL AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
DR PIR; S50999; S50999.  
DR HSP; P08223; 3SDP.  
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.  
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.  
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006801; P:superoxide metabolism; IEA.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR PRINTS; PR01703; MNSODISMUTASE.

DR ProDom: PD000475; SODismutase; 1.

KW Oxidoreductase.  
SQ SEQUENCE 49 AA; 5721 MW; C47F239DCC4E3868 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 49;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120

DB 13 ALEPYIS 19

RESULT 57

Q8XW59

ID Q8XW59 PRELIMINARY; PRT; 68 AA.

AC Q8XW59; 2.9%; Score 7; DB 2; Length 49;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DE Hypothetical protein RSC2616.  
GN RSC2616 OR RS00918.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=GM11000;  
RC MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artigienave F., Gouzy J., Mangenot S.,  
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
RL Nature 415:497-502 (2002).  
DR EMBL; AL646071; CAD16323.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 68 AA; 7745 MW; 43D0289CF7344211 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 68;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115

DB 50 PITGRAL 56

RESULT 58

Q84ER2

ID Q84ER2 PRELIMINARY; PRT; 69 AA.

AC Q84ER2; 2.9%; Score 7; DB 2; Length 69;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DE Hypothetical protein.  
OS Ralstonia solanacearum.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=96344;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=A5;  
RA Tousseaint A.C., Merlin C., Mergay M., Springael D.;  
RT "The Biphenyl Catabolic Transposon Tn4371, a Member of a New Family of  
RT Genomic Islands Related to IncP and Ti Plasmids";  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ536756; CAD61122.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 69 AA; 7844 MW; BFD6917AD4616B2 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 69;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115

DB 51 PITGRAL 57

RESULT 59

Q67050

ID Q67050 PRELIMINARY; PRT; 109 AA.

AC Q67050; 2.9%; Score 7; DB 2; Length 69;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Hemagglutinin (fragment).  
GN HA.  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=A/Fukuoka/C29/85;  
RC MEDLINE=81030852; PubMed=7421990;  
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;  
RT "cloning and dna sequence of double-stranded copies of haemagglutinin  
RT genes from h2 and h3 strains elucidates antigenic shift and drift in  
RT human influenza virus";  
RL Nature 287:301-306 (1980).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=A/Fukuoka/C29/85;  
RC MEDLINE=93233219; PubMed=7682624;  
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;  
RT "A common neutralizing epitope conserved between the hemagglutinins of  
RT influenza A virus H1 and H2 strains";  
RL J. Virol. 67:2552-2558 (1993).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; D13581; BAA02776.1; -.  
DR HSSP; P03437; 1HTM.  
DR GO; GO:0019031; C:Viral envelope; IEA.  
DR InterPro; IPR008980; Capsid hemag.  
DR InterPro; IPR001364; Hemagglutn.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
DR Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 1  
FT NON\_TER 109 109  
SQ SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGOA 171

DB 68 SEGTGOA 74

RESULT 60

Q67053

ID Q67053 PRELIMINARY; PRT; 109 AA.

AC Q67053;

DT 01-NOV-1996 (TREMELrel. 01, Created)

```

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Suita/1/90;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza a virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; D13582; BAA02777.1; -.
DR HSSP; P03437; IHGE.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW NON_TER 1
FT NON_TER 109
SQ SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 68 SEGTGQA 74
|||||

RESULT 62
Q67052 ID Q67052 PRELIMINARY; PRT; 109 AA.
AC Q67052;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA.
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ibaraki/1/90;
RX MEDLINE=81030852; PubMed=7421990;
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
RT "cloning and dna sequence of double-stranded copies of haemagglutinin
RT genes from h2 and h3 strains elucidates antigenic shift and drift in
RT human influenza virus.";
RL Nature 287:301-306(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ibaraki/1/90;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza a virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; D13583; BAA02778.1; -.
DR HSSP; P03437; IHTW.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
KW NON_TER 1
FT NON_TER 109
SQ SEQUENCE 109 AA; 12321 MW; 4B1FB7993BEC416D CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 68 SEGTGQA 74
|||||

RESULT 61
Q67051 ID Q67051 PRELIMINARY; PRT; 109 AA.
AC Q67051;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN HA.
OS Influenzavirus A.
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Sichuan/2/87;
RX MEDLINE=81030852; PubMed=7421990;
RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
RT "cloning and dna sequence of double-stranded copies of haemagglutinin
RT genes from h2 and h3 strains elucidates antigenic shift and drift in
RT human influenza virus.";
RL Nature 287:301-306(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Sichuan/2/87;
RX MEDLINE=93233219; PubMed=7682624;
RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
RT "A common neutralizing epitope conserved between the hemagglutinins of
RT influenza a virus H1 and H2 strains.";
RL J. Virol. 67:2552-2558(1993).

```

```
FT NON TER 109 109
SQ SEQUENCE 109 AA; 12347 MW; 913715121B8F52C6 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 109;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
DB 68 SEGTGQA 74
|||||

RESULT 63
Q89PD4 PRELIMINARY; PRT; 116 AA.
AC Q89PD4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BL3548 protein.
GN BLK3548.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005948; BAC48813.1; -.
KW Complete proteome.
SQ SEQUENCE 116 AA; 13293 MW; 21E7C04C4D069A31 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 116;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 EEFRASP 186
DB 23 EEFRASP 29
|||||

RESULT 64
Q8EKD6 PRELIMINARY; PRT; 118 AA.
AC Q8EKD6;
DT 01-VAR-2003 (TrEMBLrel. 23, Created)
DT 01-VAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-VAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN SO0159.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouiri H., Gill J., Utterback T.R., McDonald L.A.,
```

```
RA Felblyum T.V., Smith H.O., Vetter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015466; AAN53246.1; -.
DR TIGR; SO0159; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 118 AA; 13685 MW; 97D1F48EEEC1B45 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 118;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SRNDYSY 95
DB 70 SRNDYSY 76
|||||

RESULT 65
Q9YEN1 PRELIMINARY; PRT; 122 AA.
AC Q9YEN1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0547.
GN APE0547.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KI.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79515.1; -.
DR PIR; C72639; C72639.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 122 AA; 13447 MW; 0BA5685A116D977B CRC64;

Query Match 2.9%; Score 7; DB 17; Length 122;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 DLGTGLS 66
DB 52 DLGTGLS 58
|||||

RESULT 66
Q68196 PRELIMINARY; PRT; 125 AA.
AC Q68196;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Diol dehydratase-reactivating factor small subunit.
GN DDB.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN=ATCC 8724;
RX MEDLINE=98070363; PubMed=9405397;
RA Mori K., Tobimatsu T., Hara T., Toraya T.;
RT "Characterization, sequencing, and expression of the genes encoding a
RT reactivating factor for glycerol-inactivated adenosylcobalamin-
RL dependent diol dehydratase.";
RL J. Biol. Chem. 272:32034-32041 (1997).
DR EMBL; AF017781; AAC15872.1; -.
DR PIR; T08598; T08598.
SQ SEQUENCE 125 AA; 13620 MW; 8EC5E809C59EF74C CRC64;

Query Match      2.9%; Score 7; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAV 134
Db 8 PAIAIAV 14
|||||

RESULT 67
ID Q96VT1 PRELIMINARY; PRT; 128 AA.
AC Q96VT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ST2092.
GN ST2092.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140 (2001).
DR EMBL; AP000988; BAB67195.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 128 AA; 14502 MW; E27028A746875C94 CRC64;

Query Match      2.9%; Score 7; DB 17; Length 128;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LKGRGD 8
Db 63 LKGRGD 69
|||||

RESULT 68
ID Q9MJR3 PRELIMINARY; PRT; 157 AA.
AC Q9MJR3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 1 (EC 1.6.5.3) (NADH-ubiquinone
DE oxidoreductase chain 1) (Fragment).
GN NADH1.
OS Taenia pisiformis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Taenia.

```

```

OX NCBI_TaxID=85432;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TP3;
RX MEDLINE=20416034; PubMed=10961852;
RA Gasser R.B., Zhu X., McManus D.P.;
RT "NADH dehydrogenase subunit 1 and cytochrome c oxidase subunit I
RT sequences compared for members of the genus Taenia (Cestoda).";
RL Int. J. Parasitol. 29:1965-1970 (1999).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
DR EMBL; AJ239109; CAB77257.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001694; Resp_NADH_dhl.
DR Pfam; PF00146; NADHdh; 2.
KW NAD; Oxidoreductase; Transmembrane; Ubiquinone; Mitochondrion.
FT NON_TER 1
FT NON_TER 157
SQ SEQUENCE 157 AA; 18024 MW; AED59321B1BB6AC3 CRC64;

Query Match      2.9%; Score 7; DB 8; Length 157;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 YYSNSYS 204
Db 50 YYSNSYS 56
|||||

RESULT 69
ID Q9YDQ8 PRELIMINARY; PRT; 168 AA.
AC Q9YDQ8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0859.
GN APE0859.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KJ;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101 (1999).
DR EMBL; AP000600; BAA79839.1; -.
DR PIR; G72679; G72679.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 168 AA; 18599 MW; 267B8179D83E3C6E CRC64;

Query Match      2.9%; Score 7; DB 17; Length 168;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDGSPA 13
Db 113 GDGSPA 119
|||||

```

```
RESULT 70
O97232 PRELIMINARY; PRT; 171 AA.
AC O97232;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein, PFC0205c (l-cys-glutaredoxin-like
DE protein-1).
GN PFC0205C, WAL32.10 OR GUP-1.
OS Plasmodium falciparum (isolate 3D7), and
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329, 5833;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-P falciparum (isolate 3D7); STRAIN=3D7;
RX MEDLINE=99376085; PubMed=1048855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Hassall B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrall B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-P falciparum;
RX MEDLINE=21463082; PubMed=11479312;
RA Rahlfis S., Fischer M., Becker K.;
RT "Plasmodium falciparum Possesses a Classical Glutaredoxin and a
RT Second, Glutaredoxin-like Protein with a PICOT Homology Domain.";
RL J. Biol. Chem. 276:37133-37140 (2001).
DR ENBL; ALQ34558; CAB38997.1;
DR ENBL; AV014839; AAK0581.1;
DR GO; GO:0005489; P:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002109; Glutaredoxin.
DR InterPro; IPR004480; Glutaredox-rel.
DR Pfam; PF00462; Glutaredoxin; 1.
DR TIGRFAMs; TIGR00365; TIGR00365; 1.
SQ SEQUENCE 171 AA; 19920 MW; 91E18E5E09E5E8267 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 171;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 227 GELEKII 233
Db 164 GELEKII 170

RESULT 71
O8HKQ0 PRELIMINARY; PRT; 173 AA.
AC O8HKQ0;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 6.
GN ND6.
OS Aspasma minima.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;
OC Gobiesocidae; Gobiesocidae; Aspasma.
OX NCBI_TaxID=181476;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Miya M.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RA Miya M., Takeshima H., Endo H., Ishiguro N.B., Inoue J.G., Mukai T.,
RA Satoh T.P., Yamaguchi M., Kawaguchi A., Mabuchi K., Shirai S.M.,
RA Nishida M.;
RT "Major Patterns of higher teleostean phylogenies: A new perspective
RT based on 100 complete mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 26:121-138 (2002).
DR ENBL; AP004453; BAC23798.1;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u...; IEA.
DR InterPro; IPR001457; Oxidored_q3.
DR Pfam; PF00499; oxidored_q3; 1.
DR Mitochondrion.
KW Mitochondrion.
SQ SEQUENCE 173 AA; 17968 MW; B01855C2C1F5124A CRC64;

Query Match 2.9%; Score 7; DB 8; Length 173;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 37 GTVPLYS 43
Db 132 GTVPLYS 138

RESULT 72
O82WJ0 PRELIMINARY; PRT; 174 AA.
AC O82WJ0;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Possible cytochrome-c oxidase (EC 1.9.3.1) chain II (EC 1.9.3.1).
DR COXB OR NE0684.
GN Nitrosomonas europaea.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
DR ENBL; BX321858; CAD84595.1;
DR GO; GO:0009481; P:aa3-type cytochrome c oxidase; IEA.
DR GO; GO:0009482; P:ba3-type cytochrome c oxidase; IEA.
DR GO; GO:0009483; P:caa3-type cytochrome c oxidase; IEA.
DR GO; GO:0009485; P:cb3-type cytochrome c oxidase; IEA.
DR GO; GO:0005507; P:copper ion binding; IEA.
DR GO; GO:0004129; P:cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR InterPro; IPR001505; Copper_CuA.
DR InterPro; IPR008972; Cupredoxin.
DR PROSITE; PS00078; COX2; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 174 AA; 19495 MW; 6335A898CC509CDD CRC64;

Query Match 2.9%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 221 PSTVKAG 227
Db 97 PSTVKAG 103
```

```
RESULT 73
Q8SVV5 Q8SVV5 PRELIMINARY; PRT; 175 AA.
ID Q8SVV5;
AC Q8SVV5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ECU04_0590.
GN ECU04_0590.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope; (APR-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prentier G., Barbe V., Peyretallade E., Brotter P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453 (2001).
DR EMBL: AL590444; CAD25245.1; -.
KW Hypothetical protein.
SQ SEQUENCE 175 AA; 20432 MW; 97717F195ADB7D11 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 175;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
DB 149 VPLYSGF 155
|||||

RESULT 74
Q8ND70 Q8ND70 PRELIMINARY; PRT; 182 AA.
ID Q8ND70;
AC Q8ND70;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DAF2P667B084.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph node;
RA Ausorge W., Winkner U., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL834167; CAD38869.1; -.
DR GO: GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO: GO:0006470; F:protein amino acid dephosphorylation; IEA.
DR InterPro: IPR000222; PP2C.
DR Pfam: PF00481; PP2C; 1.
DR PROSITE: PS01032; PP2C; 1.
KW Hypothetical protein.
SQ SEQUENCE 182 AA; 20239 MW; 0D2523DE99A810BB CRC64;

Query Match 2.9%; Score 7; DB 4; Length 182;
```

```
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61
|||||

RESULT 75
Q97ZW3 Q97ZW3 PRELIMINARY; PRT; 187 AA.
ID Q97ZW3;
AC Q97ZW3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SSO0461.
GN SSO0461.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-NGOC H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
DR EMBL: AE006678; AAK40785.1; -.
DR PIR: B90191; B90191.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 187 AA; 21247 MW; 8EC5517CA85D5P22 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 187;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIIS 234
DB 61 ELEKIIIS 67
|||||

RESULT 76
Q59673 Q59673 PRELIMINARY; PRT; 189 AA.
ID Q59673;
AC Q59673;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1) (Fragment).
GN SOD.
OS Propionibacterium freudenreichii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=1744;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PZ3;
RX MEDLINE=95074560; PubMed=748202;
RA Gabbianelli R., Battistoni A., Polizio F., Carri M.T., De Martino A.,
RA Meier B., Desideri A., Rotilio G.;
RT "Metal uptake of recombinant cambialistic superoxide dismutase from
RT Propionibacterium shermanii is affected by growth conditions of host
RT Escherichia coli cells.";
RL Biochem. Biophys. Res. Commun. 216:841-847 (1995).
RN [2]
RP SEQUENCE FROM N.A.
```



```

RC STRAIN=P23;
RA Gabbianelli R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC EMBL; X91650; CAA62838.1; -.
CC EMBL; Y09012; CAA70215.1; -.
DR HSSP; P80293; LAVM.
DR GO; GO:0004785; F: copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F: iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F: manganese superoxide dismutase activity; IEA.
DR GO; GO:0048872; F: metal ion binding; IEA.
DR GO; GO:0016954; F: nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006801; P: superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sofde; 1.
DR Pfam; PF02777; sofde_C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase.
FT NON TER 189
SQ SEQUENCE 189 AA; 21334 MW; 1971419C716FF651 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120
DB 15 ALPEYIS 21

RESULT 77
Q9ZGN1 PRELIMINARY; PRT; 193 AA.
AC Q9ZGN1
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase [Mn/Fe] (EC 1.15.1.1).
GN SODB.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RA Qurollo B.A., Bishop P.E., Hassan H.M.;
RT "Identification of the iron superoxide dismutase gene sodb in
AZotobacter vinelandii.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kanematsu S., Sato S., Asada K.;
RT "Cloning of Azotobacter vinelandii Fe-superoxide dismutase gene and
its anaerobic expression in E. coli cells.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC EMBL; AF077373; AAD02836.1; -.
DR EMBL; AB025798; BAA88212.1; -.
DR HSSP; P09223; 3SDP.
DR GO; GO:0004785; F: copper, zinc superoxide dismutase activity; IEA.

```

```

DR GO; GO:0008382; F: iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F: manganese superoxide dismutase activity; IEA.
DR GO; GO:0048872; F: metal ion binding; IEA.
DR GO; GO:0016954; F: nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006801; P: superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sofde; 1.
DR Pfam; PF02777; sofde_C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase.
SQ SEQUENCE 193 AA; 21379 MW; 7E879A06D46A422B CRC64;

Query Match 2.9%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120
DB 14 ALPEYIS 20

RESULT 78
Q8JRS9 PRELIMINARY; PRT; 199 AA.
AC Q8JRS9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Attachment glycoprotein (fragment).
GN G.
OS Human respiratory syncytial virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=11250;
RN [1]
RP SEQUENCE FROM N.A.
RA Kahn J.S., Martinello R.A., Chen M.D., Weibel C.;
RT "Correlation between RSV genotype and severity of illness.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF510294; AAM82048.1; -.
DR InterPro; IPR000925; Glycoprot_G.
DR Pfam; PF08002; Glycoprotein_G; 1.
FT NON TER 199
FT NON TER 199
SQ SEQUENCE 199 AA; 21724 MW; E74D3189A27DBF74 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 199;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SQTTAIP 32
DB 31 SQTTAIP 37

RESULT 79
Q9A303 PRELIMINARY; PRT; 200 AA.
AC Q9A303;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome c oxidase assembly protein, putative.
GN CC3403.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ernolova M., White O.,
RA Saizberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE006000; AAC25365.1; -.
DR PIR; A87671; A87671.
DR TIGR; CC3403; -.
DR InterPro; IPR007533; CtaG_CoxII.
DR Pfam; PF04442; CtaG_CoxII; 1.
KW Complete proteome.
SQ SEQUENCE 200 AA; 22145 MW; 433F5432810D20AB CRC64;

Query Match 2.9%; Score 7; DB 16; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PITGRAL 115
Db 101 PITGRAL 107

RESULT 80
Q28272
ID Q28272 PRELIMINARY; PRT; 202 AA.
AC Q28272;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen type IV alpha 2 chain (Fragment).
GN COL4A2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Samoyed;
RX MEDLINE=96278820; PubMed=8662866;
RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
RT of collagen type IV. Evidence from a canine model of X-linked
RT nephritis with a COL4A5 gene mutation.";
RL J. Biol. Chem. 271:13821-13828(1996).
DR EMBL; U50934; AAC48594.1; -.
DR GO; GO:005581; C:collagen; IEA.
DR GO; GO:005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR001442; Procollag4_C.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; ProcollagNC4; 1.
DR SMART; SM00111; C4; 2.
FT NON TER 1
FT NON TER 202
SQ SEQUENCE 202 AA; 22079 MW; 25A56E7642A329FC CRC64;

Query Match 2.9%; Score 7; DB 6; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 ALAVHSQ 137
Db 105 ALAVHSQ 111

RESULT 81
Q83HP6
```

```
ID Q83HP6 PRELIMINARY; PRT; 202 AA.
AC Q83HP6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1).
GN SODA OR TW473.
OS Tropheryma whippelii (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococccineae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=218496;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22495039; PubMed=12606174;
RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,
RA Dover L.G., Norbertczak H.T., Besta G.S., Quail M.A., Harris D.E.,
RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,
RA Barrrell B.G., Parkhill J., Rellman D.A.;
RT "Sequencing and analysis of the genome of the Whipple's disease
RT bacterium Tropheryma whippelii.";
RL Lancet 361:637-644(2003).
DR EMBL; BX251411; CAD67140.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe_C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD MN; 1.
DR Oxidoreductase; Complete proteome.
KW SEQUENCE 202 AA; 22900 MW; 15305AA77FD42964 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
Db 15 ALEPYIS 21

RESULT 82
Q83G14
ID Q83G14 PRELIMINARY; PRT; 202 AA.
AC Q83G14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1).
GN SODA OR TW299.
OS Tropheryma whippelii (strain Twist) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococccineae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=203267;
RN [1]
RP SEQUENCE FROM N.A.
RA Raoult D., Audic S., Robert C., Ogata H., Suhr K., Drancourt M.,
RA Clavierie J.-M.;
RT "Tropheryma whippelii illustrates the diversity of gene loss patterns
RT in small genome bacterial pathogens.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016851; AAC44396.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
```

DR GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO:0006801; P:superoxide metabolism; IEA.  
 DR InterPro: IPR001189; SODismutase.  
 DR Pfam: PF00081; sodfe; 1.  
 DR PRINTS: PF01703; MNSODISMUTASE.  
 DR ProDom: PD000475; SODismutase; 1.  
 DR ProSITE: PS00088; SOD\_MN; 1.  
 KW Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 202 AA; 22901 MW; 9738FEEB999EBD28 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120  
 DB 15 ALEPYIS 21  
 |||||

RESULT 83

Q28274 PRELIMINARY; PRT; 205 AA.

AC Q28274; 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Collagen type IV alpha 6 chain (Fragment).  
 GN COL4A6  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Samoyed;  
 RX MEDLINE=96278820; PubMed=8662866;  
 RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;  
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains  
 of collagen type IV. Evidence from a canine model of X-linked  
 nephritis with a COL4A5 gene mutation."  
 RL J. Biol. Chem. 271:13821-13828(1996).  
 DR EMBL: J50937; AAC48587.1; -.  
 DR GO:0005581; C:collagen; IEA.  
 DR GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro: IPR001442; Procollagn4\_C.  
 DR Pfam: PF01413; C4; 2.  
 DR ProDom: PD003923; ProcollagnC4; 1.  
 DR SMART: SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 205  
 SQ SEQUENCE 205 AA; 22708 MW; 84D7BDE4C1C00395 CRC64;

Query Match 2.9%; Score 7; DB 6; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AIAVHSQ 137  
 DB 107 AIAVHSQ 113  
 |||||

RESULT 84

Q96UT6 PRELIMINARY; PRT; 206 AA.

AC Q96UT6; 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Manganese-containing superoxide dismutase (EC 1.15.1.1) (Superoxide  
 dismutase [Mn/Fe]).  
 GN SOD3.  
 OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lamarre C., LeMay J.-D., Deslauriers N., Bourbonnais Y.;  
 RT "Candida albicans expresses an unusual cytoplasmic manganese-  
 containing superoxide dismutase SOD3 upon entry and during stationary  
 phase."  
 RL J. Biol. Chem. 276:20000-20006(2001).  
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
 CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
 FAMILY.

CC EMBL: AF416340; AAL08560.1; -.  
 DR GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.  
 DR GO:0008382; F:iron superoxide dismutase activity; IEA.  
 DR GO:0008383; F:manganese superoxide dismutase activity; IEA.  
 DR GO:0046872; F:metal ion binding; IEA.  
 DR GO:0016954; F:nickel superoxide dismutase activity; IEA.  
 DR GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO:0006801; P:superoxide metabolism; IEA.  
 DR InterPro: IPR001189; SODismutase.  
 DR Pfam: PF00081; sodfe; 1.  
 DR PRINTS: PF02777; sodfe.C; 1.  
 DR ProDom: PD000475; SODismutase; 1.  
 DR ProSITE: PS00088; SOD\_MN; 1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 206 AA; 22734 MW; 303PA3503417F332 CRC64;

Query Match 2.9%; Score 7; DB 3; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120  
 DB 19 ALEPYIS 25  
 |||||

RESULT 85

Q29468 PRELIMINARY; PRT; 208 AA.

AC Q29468; 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Collagen type IV alpha 4 chain (Fragment).  
 GN COL4A4.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Samoyed;  
 RX MEDLINE=96278820; PubMed=8662866;  
 RA Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;  
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains  
 of collagen type IV. Evidence from a canine model of X-linked  
 nephritis with a COL4A5 gene mutation."  
 RL J. Biol. Chem. 271:13821-13828(1996).  
 DR EMBL: U50936; AAC48586.1; -.  
 DR GO:0005581; C:collagen; IEA.  
 DR GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR InterPro: IPR001442; Procollagn4\_C.  
 DR Pfam: PF01413; C4; 2.  
 DR ProDom: PD003923; ProcollagnC4; 1.  
 DR SMART: SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 208  
 SQ SEQUENCE 208 AA; 23135 MW; 136077AB651A21FC CRC64;

Query Match 2.9%; Score 7; DB 6; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCL 180

DB 148 SPGSCL 154

RESULT 86

Q96Y84 PRELIMINARY; PRT; 211 AA.  
 AC Q96Y84;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative superoxide dismutase.  
 GN ST2283.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM 10545 / 7;  
 RX MEDLINE=2145156; PubMed=11572479;  
 RA Kawaiyayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 Crenarchaeon, Sulfolobus tokodaii strain7";  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL; AF000989; BAB67393.1; -;  
 DR GO; GO:0046872; P:metal ion binding; IEA.  
 DR GO; GO:0004784; P:superoxide dismutase activity; IEA.  
 DR GO; GO:0006801; P:superoxide metabolism; IEA.  
 DR InterPro; IPR001189; SODismutase.  
 DR Pfam; PF00081; scdfe; 1.  
 DR Pfam; PF02777; scdfe; 1.  
 DR PRINTS; PRO1703; MNSODISMUTASE.  
 DR ProDom; PD000475; SODismutase; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 211 AA; 24302 MW; 0EEAC4AA76938002 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPVIS 120

DB 21 ALEPVIS 27

RESULT 87

P95356 PRELIMINARY; PRT; 216 AA.  
 AC P95356;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE FtsE-like protein.  
 OS Neisseria gonorrhoeae.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=485;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CH811;  
 RX MEDLINE=20277473; PubMed=10819322;

RA Bernatchez S., Francis F.M., Salimnia H., Beveridge T.J., Li H.,  
 RA Dillon J.-A.R.;  
 RT "Genomic, transcriptional and phenotypic analysis of ftsE and ftsX of  
 Neisseria gonorrhoeae";  
 RL DNA Res. 7:175-81(2000).  
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; U76418; AAB36524.1; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; P:ATP binding; IEA.  
 DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; P:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR005286; IISF.  
 DR Pfam; PF00005; ABC tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00960; 3a0501s02; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transport.  
 SQ SEQUENCE 216 AA; 23906 MW; 125F8E5363167FEB CRC64;

Query Match 2.9%; Score 7; DB 2; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GDLGTL 64

DB 63 GDLGTL 69

RESULT 88

O9KLR3 PRELIMINARY; PRT; 216 AA.  
 AC O9KLR3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Cell division ATP-binding protein FtsE.  
 GN NME0007.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58";  
 RL Science 287:1809-1815(2000).  
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 (ABC TRANSPORTERS).  
 DR EMBL; AE002359; AAF40486.1; -;  
 DR PIR; E81247; E81247.  
 DR TIGR; NMB0007; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; P:ATP binding; IEA.  
 DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; P:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.

```
DR InterPro: IPR005286; IISP.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfam: TIGR00960; 3a0501s02; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
SQ SEQUENCE 216 AA; 23945 MW; 11572764FC125FEC CRC64;

Query Match      2.9%; Score 7; DB 16; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGTL 64
   |||||
DB 63 GQDLGTL 69

RESULT 89
ID Q9JW1 PRELIMINARY; PRT; 216 AA.
AC Q9JW1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE ABC transporter ATP-binding protein.
GN FTSE OR NMA0254.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2431 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagals K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22431.";
RL Nature 404:502-506(2000).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(CC "ABC TRANSPORTERS").
DR EMBL; AL162752; CAB83562.1; -.
DR PIR; B82020; B82020.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR005286; IISF.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfam: TIGR00960; 3a0501s02; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
SQ SEQUENCE 216 AA; 23961 MW; 53F6E07B8B904DE2 CRC64;

Query Match      2.9%; Score 7; DB 16; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GQDLGTL 64
   |||||
DB 63 GQDLGTL 69
```

```
RESULT 90
ID Q8MU16 PRELIMINARY; PRT; 220 AA.
AC Q8MU16;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mn superoxide dismutase (EC 1.15.1.1) (Superoxide dismutase
DE [Mn/Fel]).
OS Trichinella pseudospiralis.
OC Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
OC Trichinellidae; Trichinella.
OX NCBI_TaxID=6337;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu H.W.K., Ko R., Mak C.H.;
RT "Characterization and molecular cloning of Mn- superoxide dismutase in
RT Trichinella pseudospiralis.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
DR EMBL; AF521909; AAM76074.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe_1.
DR Pfam: PF02777; sodfe_C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR ProDom: PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase.
SQ SEQUENCE 220 AA; 24574 MW; 85BA976A1P367FF7 CRC64;

Query Match      2.9%; Score 7; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
   |||||
DB 32 ALEPYIS 38

RESULT 91
ID Q8IXG7 PRELIMINARY; PRT; 233 AA.
AC Q8IXG7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UG0882E07.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y.;
RT "Isolation of full-length cDNA clones from human fetal brain cDNA
RT library.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF351614; AAN76514.1; -.
DR PIR; PT0240; PT0240.
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
```

DR GO: GO:0003824; F: catalytic activity; IEA.  
DR GO: GO:0004722; F: protein serine/threonine phosphatase activity; IEA.  
DR GO: GO:0005470; P: protein amino acid dephosphorylation; IEA.  
DR InterPro: IPR000222; PP2C.  
DR InterPro: IPR001932; PP2C-like.  
DR Pfam: PF00481; PP2C; 1.  
DR PROSITE: PS01032; PP2C; 1.  
SQ SEQUENCE 233 AA; 25652 MW; EB90A7B3BC1BDD08 CRC64;

Query Match 2.9%; Score 7; DB 4; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
DB 55 SGSPATW 61  
|||||

RESULT 92  
QYV4W7 PRELIMINARY; PRT; 236 AA.  
ID QYV4W7  
AC QYV4W7  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN PWT1820.  
OS Prochlorococcus marinus (strain MIT 9313).  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=74547;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22825698; PubMed=12917642;  
RA Roca G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Stegling C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
RT niche differentiation."  
RL Nature 424:1042-1047(2003).  
DR EMBL; BX572100; CAE21995.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 236 AA; 27651 MW; 5964906532FEB20E CRC64;

Query Match 2.9%; Score 7; DB 16; Length 236;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 FLFCNVN 82  
DB 83 FLFCNVN 89  
|||||

RESULT 93  
ID Q9K8B8 PRELIMINARY; PRT; 237 AA.  
AC Q9K8B8  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein BH3088.  
GN BH3088.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C-125 / JCM 9153;  
RA MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,

DR Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis."  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001517; BAB06907.1; -.  
DR PIR; H84035; H84035.  
DR InterPro: IPR008535; DUF817.  
DR Pfam; PF05675; DUF817; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 237 AA; 27822 MW; 2F2EB9A29BF14B30 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 237;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45  
DB 100 VPLYSGF 106  
|||||

RESULT 94  
ID Q8F8X2 PRELIMINARY; PRT; 239 AA.  
AC Q8F8X2  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN LA0427.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
RA Ren S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE011230; AAN47626.1; -.  
DR InterPro: IPR005660; DUF344.  
DR Pfam; PF03976; DUF344; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 239 AA; 28774 MW; ECB100FBBF7900F CRC64;

Query Match 2.9%; Score 7; DB 16; Length 239;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ELEKIIIS 234  
DB 233 ELEKIIIS 239  
|||||

RESULT 95  
ID Q8ML71 PRELIMINARY; PRT; 241 AA.  
AC Q8ML71  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE CG18279-PB.  
GN CG18279.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Change M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berland J.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-F., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phourenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berland B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Seale S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Vungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003818; AAM68571.1; -  
DR FlyBase; FBgn0033835; CG18279.

SQ SEQUENCE 241 AA; 25652 MW; EAEEA311134F9E1B CRC64;  
Query Match 2.9%; Score 7; DB 5; Length 241;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 31 IPSCPEG 37  
Db 217 IPSCPEG 223  
RESULT 96  
Q8N2M3 PRELIMINARY; PRT; 244 AA.  
AC Q8N2M3  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ90120.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,  
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,  
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK074601; BAC11085.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 244 AA; 24989 MW; F11285DC202EFC43 CRC64;  
Query Match 2.9%; Score 7; DB 4; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 171 ALASPGS 177  
Db 21 ALASPGS 27  
RESULT 97  
Q8G7Z6 PRELIMINARY; PRT; 245 AA.  
AC Q8G7Z6  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Narrowly conserved hypothetical transmembrane protein.  
CN BL0088  
OS Bifidobacterium longum.  
OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;  
OC Bifidobacteriaceae; Bifidobacterium.  
OX NCBI\_TaxID=216816;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCC 2705;  
RX MEDLINE=22294977; Pubmed=12381787;  
RA Schell M.A., Karmirantzou M., Shet B., Vilanova D., Berger B.,  
RA Pessi G., Zwhalen M.-C., Desiere F., Bork P., Delley M.,  
RA Pridmore R.D., Arigoni F.;  
RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
DR EMBL; AE014623; AAN23953.1; -  
DR GO; GO:0016021; C:integral to membrane; IEA.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
SQ SEQUENCE 245 AA; 26057 MW; E6BA8DE8DB3EC48E CRC64;



Query Match 2.9%; Score 7; DB 16; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 YSYWLST 99  
 143 YSYWLST 149  
 |||||

Db

RESULT 98  
 QXEX8 PRELIMINARY; PRT; 247 AA.  
 AC Q8XEX8; (TrEMBLrel. 20, Created)  
 DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein yacF (putative cytoplasmic protein) (Hypothetical protein STY0161).  
 GN YACF OR T0145 OR STM0139 OR STY0161.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601, 602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700933;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 Burland V., Kodyanov V., Schwartz D.C., Blattner F.R.,  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC:412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium.  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,  
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 Whitehead S., Barrall B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 DR EMBL; AEO16834; AAO67877.1; -  
 DR EMBL; AEO08700; AAL19103.1; -  
 DR EMBL; AL627265; CAD01298.1; -  
 KW Hypothetical protein; complete proteome.  
 SQ SEQUENCE 247 AA; 28425 MW; E1B9826AD004B548 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212  
 |||||

Db 157 WLASLNP 163

RESULT 99

Q87BW1

ID Q87BW1 PRELIMINARY; PRT; 252 AA.

AC Q87BW1; (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE 3-deoxy-manno-octulosonate cytidyltransferase.

GN KDSB OR PD1337.

OS Xylella fastidiosa (strain Temecula / ATCC 700964).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI\_TaxID=183190;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2242331; PubMed=12533478;

RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,

RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,

RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,

RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,

RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,

RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,

RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,

RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,

RA Cunha S.G., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,

RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,

RA Civerolo E.L., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,

RA Kitajima J.P.;

RT "Comparative analyses of the complete genome sequences of Pierce's

RT disease and citrus variegated chlorosis strains of Xylella

RT fastidiosa.";

RL J. Bacteriol. 185:1018-1026(2003).

DR EMBL; AB012558; AAC29184.1; -

DR GO; GO:0016779; Fnucoyltransferase activity; IEA.

DR GO; GO:0016740; Ftransferase activity; IEA.

DR GO; GO:0009103; Polypolysaccharide biosynthesis; IEA.

DR InterPro; IPR003329; Cytidylyl trans.

DR Pfam; PF02348; CTP transf. 3; 1

KW Transferase; Nucleotidyltransferase; Complete proteome.

SQ SEQUENCE 252 AA; 27400 MW; 1D6B0A3FFE231619 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 252;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRFTTM 74  
 |||||Db 194 LQRFTTM 200  
 |||||

RESULT 100

Q81YK7

ID Q81YK7 PRELIMINARY; PRT; 254 AA.

AC Q81YK7; (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DE Membrane protein, putative.

GN BA3535.

OS Bacillus anthracis (strain Ames).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=198094;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608414; PubMed=12721629;

RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,

RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,

RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,

RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,

```

RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Platt R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thonason B., Friedlander A.M., Koshler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AB017035; AAP27298.1; -.
DR TIGR; BA3535; -.
KW Complete proteome.
SQ SEQUENCE 254 AA; 30326 MW; 5D1554DA1E41329A CRC64;

Query Match      2.9%; Score 7; DB 16; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
DB 103 VPLYSGF 109

RESULT 101
Q81AT7 PRELIMINARY; PRT; 254 AA.
AC Q81AT7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN BC3469.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman B., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haseikorn R., Feinstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AB017009; AAP10404.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 30257 MW; 0016C115197A3D1 CRC64;

Query Match      2.9%; Score 7; DB 16; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLYSGF 45
DB 103 VPLYSGF 109

RESULT 102
Q92WM8 PRELIMINARY; PRT; 255 AA.
ID Q92WM8;
AC Q92WM8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Iron-superoxide dismutase (EC 1.15.1.1) (Superoxide dismutase
DE [unrel]).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.

```

```

OX NCBI_TaxID=4530;
RN [1]
SQ SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RX MEDLINE=99208990; PubMed=10192910;
RA Kaminaka H., Morita S., Tokumoto M., Yokoyama H., Masumura T.,
RA Tanaka K.;
RT "Molecular cloning and characterization of a cDNA for an iron-
RT superoxide dismutase in rice (Oryza sativa L.).";
RL Biosci. Biotechnol. Biochem. 63:302-308(1999).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
DR EMBL; AB014056; BAA37131.1; -.
DR PIR; JG0179; JG0179.
DR HSP; P09223; 3SDP.
DR Gramene; Q9ZWM8; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe_1.
DR Pfam; PF02777; sodfe_C; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase.
SQ SEQUENCE 255 AA; 29369 MW; 2DE803FA6161B443 CRC64;

Query Match      2.9%; Score 7; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120
DB 54 ALPEYIS 60

RESULT 103
Q9PB46 PRELIMINARY; PRT; 257 AA.
ID Q9PB46;
AC Q9PB46;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 3-deoxy-nanno-octulosonate cytidyltransferase.
GN XF2299.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorzy H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

```

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.F.O., Netto L.E.S.,  
RA Nuanli A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Tenenzy M.F., Truffi D., Tsai S.M., Tshako M.H.,  
RA Vallada H., Van Sluys M.A., Vertovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
RL Nature 406:151-159(2000).  
DR EMBL; AF004041; AAP85098.1; --  
DR FR; H82575; H82575.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008690; F:3-deoxy-manno-octulosonate cytidyltransferase; IEA.  
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.  
DR InterPro; IPR003329; Cytidylyl\_trans.  
DR InterPro; IPR004538; KdsB.  
DR Pfam; PF02348; CTP\_transf\_3; 1.  
DR TIGRFAMs; TIGR00486; KdsB; 1.  
KW Complete proteome.  
SQ SEQUENCE 257 AA; 27883 MW; 6CBB08A5D78B0BC8 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRTTMM 74  
Db 199 LQRTTMM 205

RESULT 104  
Q8SM64  
ID Q8SM64 PRELIMINARY; PRT; 259 AA.  
AC Q8SM64;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Superoxide dismutase 3 (Fragment).  
OS Toxoplasma gondii.  
OG Mitochondrion.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;  
OC Toxoplasma.  
OX NCBI\_TaxID=5811;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ding M., Kwok L.Y., Krauth-Siegel L., Schlueter D., Clayton C.,  
RA Soldati D.;  
RT "Role of catalase and peroxiredoxins as defence mechanism against  
RT oxidative stress in *Toxoplasma gondii*,"  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AIZ54045; AAO72229.1; --  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0004682; F:metal ion binding; IEA.  
DR GO; GO:0004784; F:superoxide dismutase activity; IEA.  
DR GO; GO:0006801; P:superoxide metabolism; IEA.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR Pfam; PF02777; sodfe\_C; 1.  
DR PRINTS; PR01703; MNSODISMUTASE.  
DR ProDom; PD000475; SODismutase; 1.  
KW Mitochondrion.  
FT NON\_TER 259  
SQ SEQUENCE 259 AA; 23864 MW; C0C3D8D0D6F2B6E5 CRC64;

Query Match 2.9%; Score 7; DB 8; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120  
Db 63 ALPEYIS 69

RESULT 105  
Q9SNQ0  
ID Q9SNQ0 PRELIMINARY; PRT; 259 AA.  
AC Q9SNQ0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ESTs C26547 (C12563) (EC 1.15.1.1) (Superoxide dismutase  
DE [Mn/Fel])  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC  
RL clone:PO535G04,"  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
CC -! FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -! CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -! SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
DR EMBL; AP000399; BAA83577.1; --  
DR HSSP; P09223; 3SDP.  
DR Gramene; Q9SNQ0; --  
DR GO; GO:004785; F:copper, zinc superoxide dismutase activity; IEA.  
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.  
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006801; P:superoxide metabolism; IEA.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR Pfam; PF02777; sodfe\_C; 1.  
DR PRINTS; PR01703; MNSODISMUTASE.  
DR ProDom; PD000475; SODismutase; 1.  
KW Oxidoreductase.  
SQ SEQUENCE 259 AA; 29905 MW; F685EC8AC3C15A88 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALPEYIS 120  
Db 54 ALPEYIS 60

RESULT 106  
Q34535  
ID Q34535 PRELIMINARY; PRT; 264 AA.  
AC Q34535;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE YOAT.  
GN YOAT.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]

RP SEQUENCE FROM N.A.  
RA Lapidus A., Galleron N., Sorokin A., Ehrlich D.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertoletto M.G., Bessieres P., Bolotin A., Borcherdt S.,  
RA Boursier L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrai E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,  
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maule C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yaman K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256 (1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR ENBL; AF027868; AAB84424.1; -;  
DR ENBL; Z99114; CAB13767.1; -;  
DR PIR; E69897; E69897.  
DR InterPro; IPR008535; DUF817.  
DR Pfam; PF05675; DUF817; 1.  
KW Complete proteome.  
SQ SEQUENCE 264 AA; 30605 MW; D31FD23AB2E8F667 CRC64;  
Query Match 2.9%; Score 7; DB 16; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.4e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 VPLYSGF 45  
Db 103 VPLYSGF 109  
RESULT 107  
Q92C22 PRELIMINARY; PRT; 267 AA.  
ID Q92C22  
AC Q92C22;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein lin1029.  
GN LIN1029.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
CX NCBI\_TaxID=1642;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / Serovar 6a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR ENBL; AL596167; CAC96260.1; -;  
DR PIR; AD1561; AD1561.  
DR Listalist; LIN01029; -;  
DR InterPro; IPR008535; DUF817.  
DR Pfam; PF05675; DUF817; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 267 AA; 31450 MW; 7C15IC69740DB8C2 CRC64;  
Query Match 2.9%; Score 7; DB 16; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.4e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 VPLYSGF 45  
Db 103 VPLYSGF 109  
RESULT 108  
Q9Y880 PRELIMINARY; PRT; 267 AA.  
ID Q9Y880  
AC Q9Y880;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein lmo1037.  
GN LMO1037.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
CX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR ENBL; AL591977; CAC99115.1; -;  
DR PIR; AE1204; AE1204.  
DR Listalist; LMO01037; -;  
DR InterPro; IPR008535; DUF817.  
DR Pfam; PF05675; DUF817; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 267 AA; 31204 MW; EAA24A6DEB6266D0 CRC64;  
Query Match 2.9%; Score 7; DB 16; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.4e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLXSGF 45  
|||||  
Db 103 VPLXSGF 109

## RESULT 109

Q8ET45 PRELIMINARY; PRT; 267 AA.  
AC Q8ET45;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Hypothetical conserved protein.  
GN OB0418.  
OS Oceanobacillus iheyensis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
OX NCBI\_TaxID=182710;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HTB831 / DSM 14371 / JCM 11309;  
RX MEDLINE=22220767; PubMed=12235376;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT Genome sequence of Oceanobacillus iheyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments.";  
RT Nucleic Acids Res. 30:3927-3935 (2002).  
RL EMBL; AF004594; BAC12374.1; -;  
DR InterPro; IPR008535; DUF817.  
DR Pfam; PF05675; DUF817; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 267 AA; 31291 MW; A45A3EC074FC41E3 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VPLXSGF 45  
|||||  
Db 103 VPLXSGF 109

## RESULT 110

Q7V2M5 PRELIMINARY; PRT; 274 AA.  
AC Q7V2M5;  
DT 01-OCT-2003 (TReMBLrel. 25, Created)  
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Putative multidrug efflux ABC transporter.  
GN PM00450.  
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=59919;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22825698; PubMed=12917642;  
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
RA Ahlstrom N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation."  
RL Nature 424:1042-1047 (2003).  
DR EMBL; BX572091; CAE18909.1; -;  
KW Complete proteome.  
SQ SEQUENCE 274 AA; 30471 MW; 0FOAACD4806F3BB9 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 274;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212  
|||||  
Db 209 WLASLNP 215

## RESULT 111

Q9LWS3 PRELIMINARY; PRT; 282 AA.  
AC Q9LWS3;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Similar to Zantedeschia aethiopica iron superoxide dismutase (EC 1.15.1.1) (Superoxide dismutase [Mn/Fe]).  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC clone: P0541H01.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND WHICH ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE FAMILY.  
EMBL; AP001389; BAA92737.1; -;  
DR HSP; P19665; IQNN.  
DR Gramene; Q9LWS3; -;  
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.  
DR GO; GO:0008382; F:iron superoxide dismutase activity; IEA.  
DR GO; GO:0008383; F:manganese superoxide dismutase activity; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0016954; F:nickel superoxide dismutase activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006801; F:superoxide metabolism; IEA.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR PRINTS; PR01703; MNSODISMTASE.  
DR ProDom; PD000475; SODismutase; 1;  
KW Oxidoreductase.  
SQ SEQUENCE 282 AA; 32124 MW; 6E0A907B3732CA52 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 282;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120  
|||||  
Db 35 ALEPYIS 41

## RESULT 112

Q8Q083 PRELIMINARY; PRT; 285 AA.  
AC Q8Q083;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase subunit X.  
DE X.  
GN XM0254.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=Goel / G01 / ATCC BAR-199 / DSM 3647 / OCM 88;
RX  MEDLINE=22120827; PubMed=12125824;
RA  Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA  Martinez-Arias R., Henne A., Wieser A., Baeumer S., Jacobi C.,
RA  Brueggemann H., Llenard T., Christmann A., Boencke M., Steckel S.,
RA  Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA  Fritz H.-J., Gottschalk G.;
RT  "The genome of Methanosaeta mazei: evidence for lateral gene
RT  transfer between Bacteria and Archaea.";
RL  J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR  EMBL; AB013250; AM29950.1; -.
DR  GO; GO:0008168; F:mechyltransferase activity; IEA.
DR  GO; GO:0016740; F:transferase activity; IEA.
KW  Transferase; Methyltransferase; Complete proteome.
SQ  SEQUENCE 285 AA; 30743 MW; 55BCD6753ADB44C2 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 285;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 ITGRALE 116
Db 190 ITGRALE 196
|||||

RESULT 113
Q8F0K1 PRELIMINARY; PRT; 297 AA.
AC Q8F0K1
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN LA3492.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=56601 / Serogroup Icterohaemorrhagiae / Seroovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDJ databases.
DR EMBL; AB011506; AN50690.1; -.
DR InterPro; IPR001107; Band 7.
DR Pfam; PF01145; Band 7; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 297 AA; 34391 MW; FE14184912A8AE83 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 STVAGE 228
Db 48 STVAGE 54
|||||

RESULT 114
Q88GA3 PRELIMINARY; PRT; 304 AA.
AC Q88GA3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome c family protein.
GN PF3822.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=22423050; PubMed=12534463;
RA Neilson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Bearan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson M., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Uterback J., Bizzo M., Lee K., Kosack D., Moestl D., Medler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kewitz C., Eisen J., Timmis K.N., Dueterhoeft A., Tuemmier B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016788; AAN69416.1; -.
DR TIGR; P3822; -.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00034; cytochrome c; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 2.
KW Complete proteome.
SQ SEQUENCE 304 AA; 33251 MW; DSA3EEF698F1B716 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 YSYWLST 99
Db 147 YSYWLST 153
|||||

RESULT 115
Q8C7L9 PRELIMINARY; PRT; 313 AA.
AC Q8C7L9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DnaJ (Fragment).
GN DNAJG6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049935; BAC33992.1; -.
DR FIR; PT0635; PT0696.
DR MGD; MGI:1919935; Dnajc6.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DNAJ_2; 1.
FT NON_TER 1
SQ SEQUENCE 313 AA; 33978 MW; 96C5E25DB98B957F CRC64;

Query Match 2.9%; Score 7; DB 11; Length 313;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 DLGTLS 66
|||||

```

Db 81 DLGTLGS 87

## RESULT 116

Q82ZY2 PRELIMINARY; PRT; 313 AA.  
 AC Q82ZY2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Binding protein, putative.  
 GN PAE0020.  
 OS Pyrobaculum aerophilum.  
 CC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 CC Thermoproteaceae; Pyrobaculum.  
 CX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 aerophilum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL; AE009746; AAL62507.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 313 AA; 34199 MW; E754B5E72B6CA7C4 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 GTGQALA 173

Db 90 GTGQALA 96  
 |||||

## RESULT 117

Q98L58 PRELIMINARY; PRT; 316 AA.  
 AC Q98L58;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Probable mdcF malonate transporter.  
 GN MLL1169.  
 OS Rhizobium loti (Mesorhizobium loti).  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Phyllobacteriaceae; Mesorhizobium.  
 CX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shingo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002996; BAB48605.1; -;  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR InterPro; IPR004776; Auxin\_eff.  
 DR Pfam; PF03547; Auxin\_eff; I.  
 KW Complete proteome.  
 SQ SEQUENCE 316 AA; 33254 MW; 329010C7D28EA1F6 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TGOALAS 174  
 |||||  
 Db 284 TGOALAS 290  
 |||||

## RESULT 118

Q88B27 PRELIMINARY; PRT; 331 AA.  
 AC Q88B27;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN PSP00205.  
 OS Pseudomonas syringae (pv. tomato).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 CC Pseudomonadaceae; Pseudomonas.  
 CX NCBI\_TaxID=323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PC3000;  
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
 RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,  
 RA White O., Fraser C., Collmer A.;  
 RT "Complete sequence of Pseudomonas syringae.";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; ASC16856; AAO53752.1; -;  
 DR TIGR; PSP00205; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 331 AA; 36137 MW; A4DD78A55437C197 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SGFSFLF 49

Db 13 SGFSFLF 19  
 |||||

## RESULT 119

Q92T66 PRELIMINARY; PRT; 338 AA.  
 AC Q92T66;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative oxidoreductase protein.  
 GN R00114 OR.SMC04138.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 CX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rameberger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galbert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL; AL591782; CAC41501.1; -;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR000683; GFO\_IDH\_MOCA.  
 DR InterPro; IPR004104; GFO\_IDH\_MOCA\_C.



DR Pfam; PF01408; GEO\_IDH\_MocA; 1.  
DR Pfam; PF02894; GEO\_IDH\_MocA\_C; 1.  
KW Complete proteome.  
SQ SEQUENCE 338 AA; 37006 MW; EBEFBE47EACD0B99 CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 338;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 223 TVKAGEL 229  
DB 132 TVKAGEL 138  
|||||  
RESULT 120  
Q9SAC7 PRELIMINARY; PRT; 343 AA.  
AC Q9SAC7  
DT 01-MAY-2000 (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE T16B5.11 protein.  
GN T16B5.11.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=sv. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Lee J., Liu A., Li J., Kremenetskaia I., Luros J., Gonzalez A.,  
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,  
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T16B5 sequence.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007354; RAD31338.1; -.  
DR PIR; A86241; A86241.  
DR InterPro; IPR004314; DUF239.  
DR Pfam; PF03080; DUF239; 1.  
SQ SEQUENCE 343 AA; 38362 MW; 441E8F36C5D324CF CRC64;  
  
Query Match 2.9%; Score 7; DB 10; Length 343;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 34 CPEGTV 40  
DB 55 CPEGTV 61  
|||||  
RESULT 121  
Q915Y7 PRELIMINARY; PRT; 354 AA.  
AC Q915Y7  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Hypothetical protein PA0549.  
GN PA0549  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Huftagle D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004491; AAG03938.1; -.  
DR PIR; C83577; C83577.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 354 AA; 41087 MW; A5B55A1E79008F3E CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 354;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 174 SPGSCLE 180  
DB 115 SPGSCLE 121  
|||||  
RESULT 122  
O56962 PRELIMINARY; PRT; 365 AA.  
ID O56962  
AC O56962;  
DT 01-JUN-1998 (TREMELrel. 06, Created)  
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Haemagglutinin (Fragment).  
GN HA  
OS Influenza A virus (A/equine/Newmarket-Bob Champion/89(H3N8)).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=71445;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=A/equine/Newmarket-Bob Champion/89(H3N8);  
RX MEDLINE=98309086; PubMed=9645136;  
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,  
RA Robertson J.S.;  
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza  
RT virus and comparison with laboratory derived viruses.";  
RL Arch. Virol. 143:891-901(1998).  
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
DR EMBL; AJ223193; CAAL1168.1; -.  
DR HSSP; P03437; 2VIU.  
DR GO; GO:0019031; Civil envelope; IEA.  
DR InterPro; IPR008980; Capsid hemag.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT CHAIN 1  
FT NON\_TER 1  
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.  
FT CHAIN 330 >365 HA2 HAEMAGGLUTININ.  
FT NON\_TER 365 365  
SQ SEQUENCE 365 AA; 40393 MW; 59ED645AB56485B1 CRC64;  
  
Query Match 2.9%; Score 7; DB 12; Length 365;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 165 SEGTGOA 171  
DB 358 SEGTGOA 364  
|||||  
RESULT 123  
O56961

```

ID OS6961 PRELIMINARY; PRT; 365 AA.
AC
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
GN HA.
OS Influenza A virus (A/equine/Lichfield/89 (H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=71440;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A /equine/Lichfield/89 (H3N8);
RX MEDLINE=98309066; PubMed=9645196;
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ223192; CAA11167.1; -.
DR HSP; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 328 HA1 HAEMAGGLUTININ.
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >365 HA2 HAEMAGGLUTININ.
FT NON_TER 365 365
FT SEQUENCE 365 AA; 40403 MW; E47DD6931659861 CRC64;
QY 165 SEGTOQA 171
DB 358 SEGTOQA 364

Query Match 2.9%; Score 7; DB 12; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTOQA 171
DB 358 SEGTOQA 364

RESULT 124
ID Q9DL24 PRELIMINARY; PRT; 371 AA.
AC
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Haemagglutinin (Fragment).
OS Influenza A virus (A/Athens/135/99 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=147343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Athens/135/99;
RX Plakofealos E.T., Markoulatos P., Spyrou N., Vamvakopoulos N.;
RT "Molecular and phylogenetic analysis of hemagglutinin and
RT neuraminidase sequences from recent human influenza type A (H3N2)
RT viral isolates in Southern Greece.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
DR EMBL; AF315566; AAC49309.1; -.
DR HSP; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR008980; Capsid hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
DR Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 371
FT NON_TER 371 371
FT SEQUENCE 371 AA; 40966 MW; FE9BF10A738C2CA3 CRC64;
QY 165 SEGTOQA 171
DB 365 SEGTOQA 371

Query Match 2.9%; Score 7; DB 12; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTOQA 171
DB 365 SEGTOQA 371

RESULT 125
ID Q96NT4 PRELIMINARY; PRT; 372 AA.
AC
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ30116.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kituchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Negahari K., Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE PP2C FAMILY.
DR EMBL; AK054678; BAB70790.1; -.
DR GO; GO:0006287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00352; PP2Cc; 1.
DR SMART; SM00351; PP2C-SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
DR Hypothetical protein; Hydrolase; Magnesium.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE 372 AA; 40983 MW; 7065B29DC79CB93B CRC64;
QY 9 SGSPATW 15
DB 55 SGSPATW 61

Query Match 2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 126
```

```
Q8N3J5          PRELIMINARY;          PRT;      372 AA.
ID Q8N3J5
AC Q8N3J5
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DAFZP761G058. (Human)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amalgam;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834271; CAD38946.1; -
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00332; PP2C; 1.
DR SMART; SM00331; PP2C-SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
KW Hypothetical protein.
SQ SEQUENCE 372 AA; 40997 MW; 9DD37ECC0EAD3313 CRC64;

Query Match          2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 127
Q8IUZ7          PRELIMINARY;          PRT;      372 AA.
ID Q8IUZ7
AC Q8IUZ7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreeas;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037552; AAH37552.1; -
DR GO; GO:0008287; C:protein serine/threonine phosphatase complex; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000222; PP2C.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF00481; PP2C; 1.
DR SMART; SM00332; PP2C; 1.
DR SMART; SM00331; PP2C-SIG; 1.
DR PROSITE; PS01032; PP2C; 1.
KW Hypothetical protein.
SQ SEQUENCE 372 AA; 41011 MW; 29927CBB2BDD32A2 CRC64;

Query Match          2.9%; Score 7; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 128
Q8H666          PRELIMINARY;          PRT;      382 AA.
ID Q8H666
AC Q8H666
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative iron superoxide dismutase.
GN OSJNBA0019F11.12.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBA0019F11.1";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002837; BAC22204.1; -
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0004784; F:superoxide dismutase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe; 1.
DR PRINTS; PR01703; MNSODISMUTASE.
DR PRODOM; PD000475; SODismutase; 1.
SQ SEQUENCE 382 AA; 42202 MW; D0C871B0FD0E3BDE CRC64;

Query Match          2.9%; Score 7; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
DB 135 ALEPYIS 141

RESULT 129
Q8UK63          PRELIMINARY;          PRT;      384 AA.
ID Q8UK63
AC Q8UK63
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
GN H3HA.
OS Influenza A virus (A/teal/Germany/wv201r/01).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=205472;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/teal/Germany/wv01r/01;
RA Werner O., Starick E., Mueller T., Muehle R.;
RL "Characterisation of avian influenza virus isolates from wild birds
from Germany";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
```

```
DR EMBL: AJ506781; CAD44999.1; -.
DR GO: GO:0019031; C:Viral envelope; IEA.
DR InterPro: IPR008980; Capsid_hemag.
DR RA: Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Saamato S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shingo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005964; BAC53232.1; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR005625; DUF337.
DR InterPro: IPR000362; Fumarate_lyase.
DR Pfam: PF03929; DUF337; 1.
DR PROSITE: PS00163; FUMARATE_LYASES; 1.
KW Complete proteome.
SQ SEQUENCE 384 AA; 42076 MW; 459731795CA5C838 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 374 SEGTGQA 380
|||||

RESULT 130
Q7XSV3 PRELIMINARY; PRT; 389 AA.
AC Q7XSV3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBAA039K24.3 protein.
GN OSJNBAA039K24.3
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.X., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL606637; CAB01784.1; -.
SQ SEQUENCE 389 AA; 42397 MW; 554D9E0BD459A02 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CLEEFRA 184
Db 94 CLEEFRA 100
|||||

RESULT 131
Q89C31 PRELIMINARY; PRT; 398 AA.
AC Q89C31;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BIL7967 protein.
GN BIL7967.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Saamato S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shingo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005964; BAC53232.1; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR005625; DUF337.
DR InterPro: IPR000362; Fumarate_lyase.
DR Pfam: PF03929; DUF337; 1.
DR PROSITE: PS00163; FUMARATE_LYASES; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 43470 MW; E143B907A7D6EA36 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 398;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALMP 104
Db 59 STPALMP 65
|||||

RESULT 132
Q9NWF7 PRELIMINARY; PRT; 403 AA.
AC Q9NWF7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000919; BAA91424.1; -.
DR InterPro: IPR000313; PWWP_domain.
DR Pfam: PF00855; PWWP; 1.
DR PROSITE: PS50812; PWWP; 1.
KW Hypothetical protein.
SQ SEQUENCE 403 AA; 45754 MW; 1B5D2124275EF437 CRC64;

Query Match 2.9%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180
Db 328 SPGSCLE 334
|||||

RESULT 133
Q8EG37 PRELIMINARY; PRT; 407 AA.
AC Q8EG37;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
```

```
GN SOL1774.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR ENBL; A2015621; AAN54827.1; -.
DR TIGR; SOL1774; -.
DR InterPro; IPR008599; Diacid_rec.
DR Pfam; PF05651; Diacid_rec; 1.
DR KX Hypothetical protein; Complete proteome.
SQ SEQUENCE 407 AA; 45442 MW; 533DD15CE0F210C CRC64;

Query Match 2.9%; Score 7; DB 16; Length 407;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GQALSP 175
DB 323 GQALSP 329
|||||

RESULT 134
ID Q9Q0L5 PRELIMINARY; PRT; 409 AA.
AC Q9Q0L5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/North Carolina/35922/98 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=101753;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Swine/North Carolina/35922/98;
RX MEDLINE=99412408; PubMed=10482643;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Krause S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
RT in American pigs.";
RL J. Virol. 73:8851-8856 (1999).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR ENBL; AF153232; AAD51239.1; -.
DR HSP; P03437; 2VIU.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR InterPro; IPR008980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 409
SQ SEQUENCE 409 AA; 45296 MW; 9866D50F46E531F2 CRC64;
```

```
Query Match 2.9%; Score 7; DB 12; Length 409;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTOGA 171
DB 374 SEGTOGA 380
|||||

RESULT 135
ID Q93SI5 PRELIMINARY; PRT; 412 AA.
AC Q93SI5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE O-antigen acetylase WbIA.
GN WbIA.
OS Burkholderia thailandensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=57975;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99141012; PubMed=9989483;
RA Deshazer D., Brett P.J., Woods D.E.;
RT "The type II O-antigenic polysaccharide moiety of Burkholderia
RT pseudomallei lipopolysaccharide is required for serum resistance and
RT virulence.";
RL Mol. Microbiol. 30:1081-1100 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Brett P.J., Burtick M.N., Woods D.E.;
RT "WbIA Activity is Required for the 2-O-Acetylation of O-Antigens
RT Expressed by Burkholderia pseudomallei and Burkholderia
RT thailandensis.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY033501; AAK51152.1; -.
DR GO; GO:0016747; F:transferase activity, transferring groups o. . .; IEA.
DR InterPro; IPR002656; Acyl_transf_3.
DR Pfam; PF01757; Acyl_transf_3; 1.
SQ SEQUENCE 412 AA; 45630 MW; 301FDF0F7538549EF CRC64;

Query Match 2.9%; Score 7; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 SRNDYSY 95
DB 331 SRNDYSY 337
|||||

RESULT 136
ID O69122 PRELIMINARY; PRT; 412 AA.
AC O69122;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative O-antigen acetylase.
GN WbIA.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei), and
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=28450; 13373;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.pseudomallei; STRAIN=10265;
RA Deshazer D., Brett P.J., Woods D.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC SPECIES=B.mallei;
RA Burtnick M.N., Brett P.J., Woods D.E.;
RT "Physical and Molecular Characterization of Lipopolysaccharide O-
  antigens from Burkholderia mallei."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF064070; AAD05450.1; -.
DR EMBL; AY028370; AAK27397.1; -.
DR GO; GO:0016747; F:transferase activity, transferring groups c. . .; IEA.
DR InterPro; IPR002656; Acyl_transf_3.
DR Pfam; PF01757; Acyl_transf_3; 1.
SQ SEQUENCE 412 AA; 45682 MW; DDDF409C70BF9747 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SRNDYSY 95
Db 331 SRNDYSY 337
|||||

RESULT 137
Q8KN89 PRELIMINARY; PRT; 416 AA.
ID Q8KN89
AC Q8KN89;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Q8F_8.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutayavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
  aeruginosa."
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498407; AAM27645.1; -.
SQ SEQUENCE 416 AA; 46332 MW; E1661BB8F401319B CRC64;

Query Match 2.9%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ISLWKGK 155
Db 295 ISLWKGK 301
|||||

RESULT 138
Q8KN80 PRELIMINARY; PRT; 416 AA.
ID Q8KN80
AC Q8KN80;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF_8.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutayavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
  aeruginosa."
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498407; AAM27645.1; -.
SQ SEQUENCE 416 AA; 46332 MW; E1661BB8F401319B CRC64;

Query Match 2.9%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ISLWKGK 155
Db 295 ISLWKGK 301
|||||
```

```
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498414; AAM27766.1; -.
SQ SEQUENCE 416 AA; 46346 MW; D6646B03603536C9 CRC64;

Query Match 2.9%; Score 7; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ISLWKGK 155
Db 295 ISLWKGK 301
|||||

RESULT 139
Q8QJ74 PRELIMINARY; PRT; 416 AA.
ID Q8QJ74
AC Q8QJ74;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/Iowa/8548-1/98).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=101751;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A/Swine/Iowa/8548-1/98;
RX MEDLINE=99412408; PubMed=10482643;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Krauss S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
  in American pigs."
RL J. Virol. 73:8851-8856(1999).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC -!- CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AF153235; AAD51242.1; -.
DR HSP; P03437; 2VIU.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR009980; Capsid_hemag.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON TER 416
SQ SEQUENCE 416 AA; 46324 MW; 753E943B03CAB0ED CRC64;

Query Match 2.9%; Score 7; DB 12; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTCQA 171
Db 374 SEGTCQA 380
|||||

RESULT 140
Q82043 PRELIMINARY; PRT; 419 AA.
ID Q82043
AC Q82043;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemorrhagic toxin a (EC 3.4.24.1) (Fragment).
GN HT-A, ATROLYSIN A.
OS Crotalus atrox (Western diamondback rattlesnake).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Crotalus.
```

```
OX NCBI_TaxID=8730;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=venom gland;
RA Hite L.A., Jia L.-G., Bjarnason J.B., Fox J.W.;
RT "cDNA sequences for four snake venom metalloproteinases: Structure,
RT classification, and their relationship to mammalian reproductive
RT proteins.";
RL Arch. Biochem. Biophys. 0:0-0(1993).
DR EMBL; U01234; AAA0326.1; -.
DR PIR; S41607; S41607.
DR HSSP; P15167; IDTH.
DR MEROPS; M12.142; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00200; disintegrin_1.
DR Pfam; PF01421; Reprolysin_1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SMC0608; ACR; 1.
DR SMART; SMC0050; DISIN; 1.
DR PROSITE; PS02015; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS02014; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase.
FT NON_TER
SQ SEQUENCE 419 AA; 46879 MW; 442833518478E416 CRC64;

Query Match 2.9%; Score 7; DB 13; Length 419;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 LFCNVND 83
Db 367 LFCNVND 373

RESULT 141
Q9GSD3 ID Q9GSD3 PRELIMINARY; PRT; 424 AA.
AC Q9GSD3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative transporter protein CG10.
GN CG10.
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=126793;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Salvador I;
RX MEDLINE=21240730; PubMed=11343215;
RA Nomura T., Carlton J.M.R., Baird J.K., Del Portillo H.A.,
RA Fryauff D.J., Rathore D., Fidock D.A., Su X.-Z., Collins W.E.,
RA McCutchan T.F., Wootton J.C., Wellem T.E.;
RT "Evidence for different mechanisms of chloroquine resistance in 2
RT Plasmodium species that cause human malaria.";
RL J. Infect. Dis. 183:1653-1661 (2001).
DR EMBL; AF314649; AAG27739.1; -.
DR EMBL; AF314646; AAG27735.1; -.
SQ SEQUENCE 424 AA; 48569 MW; EB54A287C401BFD9 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPATAIA 133
Db 352 GPATAIA 358

RESULT 142
Q9T9R7 ID Q9T9R7 PRELIMINARY; PRT; 424 AA.
AC Q9T9R7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM6;
RA Li G.-D., Ward S.A.;
RT "Plasmodium falciparum TM6 putative chloroquine resistance transporter
RT (crt) mRNA.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF468066; AAL75580.1; -.
DR EMBL; AF468066; AAL75580.1; -.
SQ SEQUENCE 424 AA; 48581 MW; DE470716070B49D6 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPATAIA 133
Db 352 GPATAIA 359

RESULT 143
Q9GSD7 ID Q9GSD7 PRELIMINARY; PRT; 424 AA.
AC Q9GSD7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative transporter protein CG10.
GN CG10.
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21240730; PubMed=11343215;
RA Nomura T., Carlton J.M.R., Baird J.K., Del Portillo H.A.,
RA Fryauff D.J., Rathore D., Fidock D.A., Su X.-Z., Collins W.E.,
RA McCutchan T.F., Wootton J.C., Wellem T.E.;
RT "Evidence for different mechanisms of chloroquine resistance in 2
RT Plasmodium species that cause human malaria.";
RL J. Infect. Dis. 183:1653-1661 (2001).
DR EMBL; AF314646; AAG27735.1; -.
DR EMBL; AF314646; AAG27735.1; -.
SQ SEQUENCE 424 AA; 48672 MW; E3F2872D08988753 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPATAIA 133
Db 352 GPATAIA 358

RESULT 144
Q9NH61 ID Q9NH61 PRELIMINARY; PRT; 424 AA.
AC Q9NH61;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```



```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
GN CRT.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21000495; PubMed=11090624;
RX Fideck D.A., Nomura T., Talley A.K., Cooper R.A., Dzekunov S.M.,
RA Ferdig M.T., Ursos L.M.B., Bir Singh Sidhu A., Naude B., Deitsch K.W.,
RA Su X.Z., Wootton J.C., Roepke P.D., Wellens T.E.;
RT "Mutations in the P. falciparum digestive vacuole transmembrane
RT protein PfCRT and evidence for their role in chloroquine resistance.";
RL Mol. Cell. 6:861-871(2000).
DR EMBL; AF233065; AAF60272.1; -.
SQ SEQUENCE 424 AA; 48619 MW; 710AE2730717AEDE CRC64;

Query Match      2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPAIAIA 133
Db 353 GPAIAIA 359

RESULT 145
Q9N623 PRELIMINARY; PRT; 424 AA.
ID Q9N623
AC Q9N623;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
GN CRT.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NF54;
RA Fideck D.A., Nomura T., Talley A.K., Cooper R.A., Dzekunov S.M.,
RA Ferdig M.T., Ursos L.M., Su X., Wootton J.C., Roepke P.D.,
RA Wellens T.E.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF233068; AAF60275.1; -.
DR EMBL; AF233066; AAF60273.1; -.
SQ SEQUENCE 424 AA; 48675 MW; EC642763C5C73732 CRC64;

Query Match      2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPAIAIA 133
Db 353 GPAIAIA 359

RESULT 146
Q8IB29 PRELIMINARY; PRT; 424 AA.
ID Q8IB29
AC Q8IB29;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chloroquine resistance transporter, putative.
GN MAL7p1.27.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson D.J., Mungthin M., Bray P.G., Ward S.A.;
RT "Identification of a novel mutation in pfCRT that confers sensitivity
RT to CQ after selective drug pressure with amantadine and
RT halofantrine.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF495378; AAO85508.1; -.
SQ SEQUENCE 424 AA; 48647 MW; 1ACE791F8745FEFC CRC64;

Query Match      2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPAIAIA 133
Db 353 GPAIAIA 359

RESULT 147
Q86W70 PRELIMINARY; PRT; 424 AA.
ID Q86W70
AC Q86W70;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
GN CRT.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson D.J., Mungthin M., Bray P.G., Ward S.A.;
RT "Identification of a novel mutation in pfCRT that confers sensitivity
RT to CQ after selective drug pressure with amantadine and
RT halofantrine.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF495376; AAO85506.1; -.
SQ SEQUENCE 424 AA; 48592 MW; 1DA58CC249C0A004 CRC64;

Query Match      2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPAIAIA 133
Db 353 GPAIAIA 359

RESULT 148
Q86W68 PRELIMINARY; PRT; 424 AA.
ID Q86W68
AC Q86W68;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative chloroquine resistance transporter.
GN CRT.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson D.J., Mungthin M., Bray P.G., Ward S.A.;
RT "Identification of a novel mutation in pfCRT that confers sensitivity
RT to CQ after selective drug pressure with amantadine and
RT halofantrine.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF495378; AAO85508.1; -.
SQ SEQUENCE 424 AA; 48647 MW; 1ACE791F8745FEFC CRC64;

Query Match      2.9%; Score 7; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GPAIAIA 133
Db 353 GPAIAIA 359

```

```
QY 127 GPALAI 133
DB 353 GPALAI 359

RESULT 149
Q9ST79 PRELIMINARY; PRT; 425 AA.
AC Q9ST79,
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CAA303718.1 protein.
GN Q3037.18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DNA;
RA Hong G., Chen Z.;
RT "Oryza sativa chromosome 4 BAC Q3037-207F1 genomic sequence.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ245900; CAB53491.1; -.
DR Gramene; Q9ST79; -.
DR InterPro; IPR008511; DUF793.
DR Pfam; PF05633; DUF793; 1.
SQ SEQUENCE 425 AA; 46455 MW; C6D3EBC56DC4B777 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 425;
Best Local Similarity 100.0%; Pred.No. 2.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CLEPRA 184
DB 94 CLEPRA 100

RESULT 150
Q9Q0L4 PRELIMINARY; PRT; 429 AA.
AC Q9Q0L4,
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hemagglutinin (Fragment).
OS Influenza A virus (A/Swine/Texas/4199-2/98(H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=136487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Swine/Texas/4199-2/98;
RA Zhou N.N., Senne D.A., Landgraf J.S., Swenson S.L., Erickson G.,
RA Rossow K., Liu L., Yoon K.J., Kraus S., Webster R.G.;
RT "Genetic reassortment of avian, swine, and human influenza A viruses
in American pigs.";
RL J. Virol. 73:8851-8856(1999).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AF153233; AADS1240.1; -.
DR HSP; P03437; 1HTW.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR InterPro; IPR008980; Capsid_hemag.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTININ2.
```

DR ProDom: PD000225; Hemagglutin; 1.  
KW Envelope protein; Glycoprotein; Hemagglutinin.  
FT NON\_TER 429 429  
SQ SEQUENCE 429 AA; 47838 MW; 72DC7925F8D02285 CRC64;

Query Match 2.9%; Score 7; DB 12; Length 429;  
Best Local Similarity 100.0%; Pred.No. 2.le+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTCQA 171  
DB 374 SEGTCQA 380

Search completed: April 5, 2004, 07:38:57  
Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:05:43 ; Search time 60 Seconds  
(without alignments)  
1149.027 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRGDSGPATWTRGF.....KAGLEKIIISRCQVCMKKEH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :

A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	100.0	244	5	ABG79219 Human Goo
2	244	100.0	244	5	Aau75595 Human typ
3	244	100.0	244	6	ADA20225 Human typ
4	244	100.0	245	3	Aay67942 Human typ
5	244	100.0	245	5	Aau75589 Human typ
6	244	100.0	1670	7	ADA47063 Human Pro
7	235	96.3	244	5	ABG79218 Human typ
8	235	96.3	244	5	ABG79217 Human typ
9	191	78.3	191	5	Aau75596 Human typ
10	191	78.3	191	6	ADA20260 Human typ
11	183	75.0	254	5	Aau75598 Human typ
12	163	66.8	268	2	Aay31993 Type IV c
13	163	66.8	268	3	Aay75555 Human alp
14	159	65.2	211	3	Aay95918 Human Goo
15	159	65.2	211	5	ABG79208 Human GP
16	155	63.5	232	7	ADCl17697 Human typ
17	141	57.8	218	2	Aay44172 Human typ
18	141	57.8	218	3	Aay56784 Human alp
19	141	57.8	218	4	Aae09484 Human alp
20	132	54.1	132	6	ADA20261 Human typ
21	131	53.7	132	5	Aau75597 Human typ
22	124	50.8	124	5	Aau75594 Human typ
23	124	50.8	124	6	ADA20258 Human typ
24	120	49.2	120	6	ADA20259 Human typ
25	112	45.9	112	6	ADA20262 Human typ

26	99	40.6	218	2	AAU79164
27	88	36.1	88	5	AAU75607
28	88	36.1	88	6	ADA20271
29	81	33.2	88	5	AAU75608
30	80	32.8	88	6	ADA20272
31	79	32.4	79	5	AAU75600
32	79	32.4	79	6	ADA20264
33	65	26.6	65	5	AAU75599
34	64	26.2	64	6	ADA20263
35	61	25.0	68	3	AAU95920
36	61	25.0	68	5	ABG79210
37	61	25.0	72	3	AAU95919
38	61	25.0	72	3	AAU95921
39	61	25.0	72	5	ABG79209
40	61	25.0	72	5	ABG79211
41	39	16.0	230	7	ADA47061
42	39	16.0	471	2	AAU44171
43	39	16.0	471	3	AAU56783
44	39	16.0	471	4	AAE09483
45	37	15.2	72	5	ABG79213
46	36	14.8	36	4	AAE09503
47	30	12.3	471	2	AAU79163
48	26	10.7	26	4	AAE09501
49	26	10.7	27	6	ADA20238
50	25	10.2	25	6	ADA20236
51	22	9.0	22	7	ADCl17661
52	22	9.0	22	7	ADCl17414
53	21	8.6	21	3	AAU95912
54	21	8.6	21	5	ABG79202
55	21	8.6	21	5	ADCl17642
56	20	8.2	20	5	AAU75604
57	20	8.2	20	5	AAU75602
58	20	8.2	20	5	AAU75603
59	20	8.2	20	6	ADA20267
60	20	8.2	20	6	ADA20266
61	20	8.2	20	6	ADA20268
62	20	8.2	20	7	ADCl17684
63	19	7.8	19	5	AAU75606
64	19	7.8	19	5	AAU75605
65	19	7.8	19	6	ABP58053
66	19	7.8	19	6	ADA20269
67	19	7.8	19	6	ADA20265
68	19	7.8	19	6	ADA20270
69	19	7.8	19	7	AAU75601
70	18	7.4	18	7	ADCl17655
71	18	7.4	18	7	ADCl17672
72	18	7.4	18	7	ADCl17649
73	17	7.0	46	4	AAU18657
74	17	7.0	46	5	ABG40558
75	17	7.0	229	1	AAU95524
76	17	7.0	229	3	AAU67943
77	17	7.0	229	5	AAU75587
78	17	7.0	229	6	ADA20217
79	17	7.0	229	7	ADCl17695
80	17	7.0	229	7	ADCl17699
81	17	7.0	260	2	AAU31991
82	17	7.0	260	3	AAU97553
83	17	7.0	264	2	AAU31995
84	17	7.0	264	3	AAU97557
85	17	7.0	309	3	AAU54044
86	17	7.0	406	3	AAU58169
87	17	7.0	772	2	AAU23873
88	17	7.0	772	2	AAU09643
89	17	7.0	1669	5	AAU40863
90	17	7.0	1669	5	ABG90760
91	17	7.0	1669	6	ABU54467
92	17	7.0	1669	6	ABU54467
93	17	7.0	1672	4	AAU39077
94	17	7.0	1685	4	ABG04839
95	17	7.0	1693	4	ABG15619
96	16	6.6	16	7	ADCl17470
97	15	6.1	15	4	AAE09491
98	15	6.1	15	4	AAE09497

AAU79164	Partial s
Aau75607	Human typ
ADA20271	Human typ
Aau75608	Human typ
ADA20272	Human typ
Aau75600	Human typ
ADA20264	Human typ
Aau75599	Human typ
ADA20263	Human typ
AAU95920	Human Goo
ABG79210	Human GP
AAU95919	Human Goo
AAU95921	Human Goo
ABG79209	Human GP
ABG79211	Human GP
ADA47061	Rat Prote
AAU44171	Bovine ty
AAU56783	Bovine al
AAE09483	Bovine al
ABG79213	Human GP
AAE09503	Human C8
AAU79163	Partial s
AAE09501	Human C7
ADA20238	T8 peptid
ADA20236	T7 peptid
ADCl17661	Type IV c
ADCl17414	Type IV c
AAU95912	Human Goo
ABG79202	Human Goo
ADCl17642	Type IV c
AAU75604	Human typ
AAU75602	Human typ
AAU75603	Human typ
ADA20267	Human typ
ADA20266	Human typ
ADA20268	Human typ
ADCl17684	Type IV c
AAU75606	Human typ
AAU75605	Human typ
ABP58053	Collagen
ADA20269	Human typ
ADA20265	Human typ
ADA20270	Human typ
AAU75601	Human typ
ADCl17655	Type IV c
ADCl17672	Type IV c
ADCl17649	Type IV c
AAU18657	Peptide #
ABG40558	Human pep
AAU95524	Complete
AAU67943	Human typ
AAU75587	Human typ
ADA20217	Human typ
ADCl17695	Human typ
ADCl17699	Human typ
AAU31991	Type IV c
AAU97553	Human alp
AAU31995	Type IV c
AAU97557	Human alp
AAU54044	Human pan
AAU58169	Lung canc
AAU23873	Human alp
AAU09643	Human typ
AAU40863	Human pol
ABG90760	Human Tum
ABU54467	Mouse isc
ABU54467	Human tum
AAU39077	Human pol
ABG04839	Novel hum
ABG15619	Novel hum
ADCl17470	Type IV c
AAE09491	Human C2
AAE09497	Human C5

99	15	6.1	15	4	AAE09499	Human C6	172	8	3.3	228	7	ADC17700	ADC17700 Human typ
100	15	6.1	15	4	AAE09495	Human C4	173	8	3.3	260	3	AAI31996	AAI31996 Type IV c
101	15	6.1	15	7	ADC17449	Type IV c	174	8	3.3	260	3	AAI31996	AAI31996 Type IV c
102	15	6.1	15	7	ADC17586	Type IV c	175	8	3.3	1779	4	ABE60207	ABE60207 Drosophil
103	15	6.1	15	7	ADC17607	Type IV c	176	7	2.9	7	6	ADA20248	ADA20248 Peptide S
104	15	6.1	15	7	ADC17626	Type IV c	177	7	2.9	7	6	ADA20254	ADA20254 Peptide S
105	15	6.1	15	7	ADC17532	Type IV c	178	7	2.9	7	6	ADA20254	ADA20254 Peptide S
106	15	6.1	15	7	ADC17490	Type IV c	179	7	2.9	13	2	AAW94939	AAW94939 MHC bindi
107	14	5.7	14	7	ADC17560	Type IV c	180	7	2.9	14	2	AAE09494	AAE09494 Human C4
108	14	5.7	14	7	ADC17427	Type IV c	181	7	2.9	14	7	ADC17425	ADC17425 Type IV c
109	14	5.7	14	7	ADC17508	Type IV c	182	7	2.9	15	7	ADC17488	ADC17488 Type IV c
110	14	5.7	14	7	ADC17402	Type IV c	183	7	2.9	15	7	ADC17625	ADC17625 Type IV c
111	14	5.7	14	7	ADC17667	Type IV c	184	7	2.9	15	7	ADC17627	ADC17627 Type IV c
112	14	5.7	14	7	ADC17605	Type IV c	185	7	2.9	18	3	AAI79501	AAI79501 Bimeria t
113	14	5.7	14	3	AAE09336	Human epi	186	7	2.9	35	4	AAE09502	AAE09502 Human C8
114	14	5.7	14	3	AAE095913	Human Goo	187	7	2.9	70	4	AAE17803	AAE17803 Peptide #
115	14	5.7	21	5	ABG79203	Human Goo	188	7	2.9	70	4	AAE17803	AAE17803 Peptide #
116	14	5.7	25	6	ADA20237	T7 mutant	189	7	2.9	70	4	ABE68827	ABE68827 Peptide #
117	12	4.9	12	2	AAI44173	Bovine ty	190	7	2.9	70	4	AAE030312	AAE030312 Peptide #
118	12	4.9	12	3	AAI56785	Human alp	191	7	2.9	70	4	ABE31608	ABE31608 Peptide #
119	12	4.9	12	4	AAE09485	Human alp	192	7	2.9	70	4	ABE22150	ABE22150 Protein #
120	12	4.9	12	4	AAE09493	Human C3	193	7	2.9	70	4	AAE69972	AAE69972 Human bon
121	12	4.9	12	4	AAE97334	Collagen	194	7	2.9	70	4	AAE57569	AAE57569 Human bra
122	12	4.9	12	4	AAE97335	Collagen	195	7	2.9	70	4	ABG51672	ABG51672 Human liv
123	12	4.9	12	4	AAE97333	Collagen	196	7	2.9	70	4	AAE03881	AAE03881 Peptide #
124	12	4.9	12	6	ABP58086	Peptide u	197	7	2.9	70	4	AAE05450	AAE05450 Peptide #
125	12	4.9	12	7	ADC17409	Type IV c	198	7	2.9	70	5	ABG39604	ABG39604 Human pep
126	11	4.5	11	7	ADC17634	Type IV c	199	7	2.9	70	5	ABG37889	ABG37889 Human pep
127	11	4.5	11	7	ADC17571	Type IV c	200	7	2.9	79	4	AAE90264	AAE90264 Human imm
128	11	4.5	11	7	ADC17545	Type IV c							
129	11	4.5	11	7	ADC17666	Type IV c							
130	11	4.5	11	6	ADA20240	TP3 pepti							
131	11	4.5	22	7	ADC17413	Type IV c							
132	11	4.5	22	7	ADC17415	Type IV c							
133	11	4.5	1940	4	ABE64070	Drosophil							
134	10	4.1	10	4	AAE93938	Human com							
135	10	4.1	10	4	AAE97338	Collagen							
136	10	4.1	10	4	AAE97336	Collagen							
137	10	4.1	10	6	ADA20257	Peptide S							
138	10	4.1	10	7	ADC17678	Type IV c							
139	10	4.1	10	7	ADC17691	Type IV c							
140	10	4.1	27	6	ADA20239	T8-3 pept							
141	10	4.1	27	6	ADA20241	P2 peptid							
142	10	4.1	143	5	ABG70168	Human pre							
143	10	4.1	227	6	ADA20221	Human typ							
144	10	4.1	227	6	ADC17696	Human typ							
145	10	4.1	241	5	AAU75588	Human typ							
146	10	4.1	241	5	AAU75588	Human typ							
147	10	4.1	242	3	AAI57946	Human typ							
148	10	4.1	258	2	AAI31992	Type IV c							
149	10	4.1	258	3	AAI31992	Type IV c							
150	10	4.1	430	3	AAE58180	Lung canc							
151	10	4.1	430	3	AAE58180	Lung canc							
152	10	4.1	459	5	ABP59082	Human pol							
153	10	4.1	900	5	ABP43610	Alpha-2 t							
154	9	3.7	1712	5	AAI17361	Human alp							
155	9	3.7	12	6	ADA20256	Peptide S							
156	9	3.7	12	7	ADC17408	Type IV c							
157	9	3.7	16	7	ADC17472	Type IV c							
158	9	3.7	18	7	ADC17654	Type IV c							
159	9	3.7	18	7	ADC17656	Type IV c							
160	9	3.7	22	7	ADC17660	Type IV c							
161	8	3.3	8	4	AAE97337	Collagen							
162	8	3.3	8	4	AAE97337	Collagen							
163	8	3.3	8	6	ABP58055	Peptide u							
164	8	3.3	8	6	ADA20249	Peptide S							
165	8	3.3	8	6	ADA20255	Peptide S							
166	8	3.3	14	3	AAI55917	Human Goo							
167	8	3.3	14	5	AAE95917	Human Goo							
168	8	3.3	18	7	ADC17648	Human typ							
169	8	3.3	18	7	ADC17648	Human typ							
170	8	3.3	19	3	AAI55916	Human Goo							
171	8	3.3	19	5	ABG79206	Human typ							
			20	7	ADC17683	Type IV c							

## ALIGNMENTS

## RESULT 1

ABG79219  
 ID ABG79219 standard; protein; 244 AA.  
 XX  
 AC ABG79219;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human Goodpasture disease-related protein.  
 XX  
 KW Goodpasture antigen binding protein; Goodpasture syndrome;  
 KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
 KW autoimmune condition; phosphorylation; myelin basic protein; MBP;  
 KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;  
 KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
 KW pemphigoid; lichen planus.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS  
 PN W0200261430-AA.  
 PD 08-AUG-2002.  
 PD  
 PF 31-JAN-2002; 2002WO-EP001010.  
 PF  
 PR 31-JAN-2001; 2001US-0265249P.  
 PR  
 PA (SAUS/) SAUS J.  
 PA  
 PI SAUS J;  
 PI  
 DR WPI; 2002-619280/66.  
 DR  
 XX  
 XX  
 PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
 PT Goodpasture syndrome or lupus, comprises identifying compounds that  
 PT reduce phosphorylation of, or formation of conformational isomers of,  
 PT target proteins.  
 PT  
 XX

PS Disclosure; Page 213-214; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an

CC autoimmune condition by identifying compounds that reduce phosphorylation

CC of a first target protein (I) (which is selected from Goodpasture antigen

CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)

CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-

CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising

CC Glu-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of

CC conformational isomers of the second target protein (II) (selected from

CC an alpha3 type IV collagen NC1 domain polypeptide and myelin basic

CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3

CC NC1 domain conformational isomer, which has an amino acid sequence

CC identical to the wild type alpha3 type IV collagen NC1 domain, is

CC stabilised by disulphide bonds, and has a molecular weight in a non-

CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in

CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated

CC type IV collagen alpha3 NC1 domain. The human gene for GPBP is located on

CC chromosome 5q13. The method is useful for treating autoimmune conditions,

CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous

CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present

CC sequence represents a Goodpasture syndrome related protein or peptide

XX

SQ Sequence 244 AA;

Query Match 100.0%; Score 244; DB 5; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKXRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSGFSLVQGNQRAHQD 60

DB 1 GLKXRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSGFSLVQGNQRAHQD 60

QY 61 LGTGLSCLORFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120

DB 61 LGTGLSCLORFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

QY 161 EFRASPFLECHGRGTCNYNSYSFSLASLNPERFRKPISTVKAGELEKIIISRCQVCM 240

DB 161 EFRASPFLECHGRGTCNYNSYSFSLASLNPERFRKPISTVKAGELEKIIISRCQVCM 240

QY 241 KKRH 244

DB 241 KKRH 244

RESULT 2

ID AAU75595

XX AAU75595 standard; protein; 244 AA.

AC AAU75595;

XX

DT 08-MAY-2002 (first entry)

XX

DE Human type IV collagen alpha 3 chain mutant, Tumstatin 334.

XX

KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;

KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;

KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;

XX Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

XX

PN WO200151523-A2.

XX

PD 19-JUL-2001.

XX

PF 08-JAN-2001; 2001WO-US000565.

XX

PR 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-198037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and

PT treating disorders involving angiogenesis.

XX Example 33; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1

CC domain, having one or more of the characteristics selected from: (a) the

CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause

CC apoptosis of endothelial cells. Also described are the following: (1) use

CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for

CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,

CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by

CC promoting or inducing endothelial cell apoptosis in a tissue, where the

CC endothelial cell apoptosis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; (2) use of

CC an antibody or peptide that specifically binds the alpha1, alpha2, or

CC alpha3, alpha4, alpha5, alpha6, beta1 or beta3 subunit of integrin in the

CC preparation of a medicament for inhibiting angiogenesis or cell

CC proliferation; (3) use of an inhibitor, such as an antibody, antibody

CC fragment or peptide of receptor-mediated angiogenesis in a vertebrate,

CC where the disease is characterised by angiogenesis that is mediated by

CC receptors to Arresten, Canstatin or Tumstatin and where the receptors

CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one

CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in

CC the presence of a medicament for promoting angiogenesis in a tissue; and

CC (5) use of integrins in the preparation of a medicament for promoting or

CC inducing angiogenesis or cell proliferation in a tissue. The fragments

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues

CC or allelic variants are useful in the preparation of a medicament for

CC treating a disorder involving inhibiting angiogenesis in a tissue, where

CC the angiogenesis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits; or by promoting or

CC inducing endothelial cell apoptosis in a tissue, where the endothelial

CC cell apoptosis is mediated by one or more endothelial cell integrins or

CC one or more endothelial cell integrin subunits. The medicament is useful

CC in inhibiting tumour growth and for the regression of an established

CC tumour. The present sequence represents the amino acid sequence of human

CC type IV collagen alpha 3 chain mutant, Tumstatin 334, which consists of

CC residues 2-245 of Tumstatin. Note: The present sequence is not shown in

CC the specification but is derived from the wild type human Tumstatin

CC sequence given in figure 18A (see AAU75595)

XX

SQ Sequence 244 AA;

Query Match 100.0%; Score 244; DB 5; Length 244;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKXRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSGFSLVQGNQRAHQD 60

DB 1 GLKXRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSGFSLVQGNQRAHQD 60

QY 61 LGTGLSCLORFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120

DB 61 LGTGLSCLORFTHMPFLFCNVNDVCFASRNDYSYWLSTPALMPNMNAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSTFIMFTSAGSEGTQALASPGSCLE 180

QY 181 EFRASPFLCHGRCGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCM 240  
 Db 181 EFRASPFLCHGRCGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCM 240  
 QY 241 KKH 244  
 Db 241 KKH 244

RESULT 3  
 ADA20225  
 ID ADA20225 standard; protein; 244 AA.  
 XX AC ADA20225;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Human type IV collagen alpha 3 chain partial protein sequence.  
 XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human.  
 XX OS Homo sapiens.  
 XX PH Key  
 FT Region 1. .244 Location/Qualifiers  
 FT /note= "Tumstatin"  
 FT Region 1. .124  
 FT /note= "Tumstatin 333; pET22b-alpha 3 (IV) NC1 region"  
 FT Region 1. .19  
 FT /note= "T1 peptide"  
 FT Region 28. .42  
 FT /note= "First Goodpasture epitope"  
 FT Region 53. .72  
 FT /note= "T2 peptide"  
 FT Region 58. .88  
 FT /note= "T3 peptide"  
 FT Region 83. .102  
 FT /note= "T4 peptide"  
 FT Region 98. .116  
 FT /note= "T5 peptide"  
 FT Region 113. .132  
 FT /note= "T6 peptide"  
 FT Region 125. .244  
 FT /note= "Tumstatin 334"  
 FT Region 139. .152  
 FT /note= "Second Goodpasture epitope"  
 XX PN WO200305257-A2.  
 XX PD 24-JUL-2003.  
 XX PF 20-DEC-2002; 2002WO-US040938.  
 XX PR 21-DEC-2001; 2001US-00032221.  
 XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX PI Kalluri R;  
 XX PF 2003-587256/55.  
 XX DR N-PSDB; ADA20224.  
 XX PF New peptide, useful for preparing a composition for inhibiting tumor  
 XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX PS Claim 52; Fig 18; 240pp; English.  
 XX CC This invention relates to novel isolated proteins and their fragments

CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the partial amino acid sequence of the alpha 3 chain of human  
 CC type IV collagen. The "tumstatin" protein of the invention was derived  
 CC from this protein and comprises the full length of the present sequence.  
 XX SQ Sequence 244 AA;  
 Query Match 100.0%; Score 244; DB 6; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-241;  
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GLKKGKDGSGSPATWTRGTFVTRHSQTTAIPSCBEGTVPLYSFSLFVGNORAHGQD 60  
 Db 1 GLKKGKDGSGSPATWTRGTFVTRHSQTTAIPSCBEGTVPLYSFSLFVGNORAHGQD 60  
 QY 61 LGTLGSLQRFMTTPELFCNVNDVCFASRDNDYSYWLSTPALMPNMAPITGRALEPYIS 120  
 Db 61 LGTLGSLQRFMTTPELFCNVNDVCFASRDNDYSYWLSTPALMPNMAPITGRALEPYIS 120  
 QY 121 RCTVCGSPATAVAHSQTTDIPCPHGWSLWKGFSTMTSAGSEGTGQALASPGSCLE 180  
 Db 121 RCTVCGSPATAVAHSQTTDIPCPHGWSLWKGFSTMTSAGSEGTGQALASPGSCLE 180  
 QY 181 EFRASPFLCHGRCGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCM 240  
 Db 181 EFRASPFLCHGRCGTCNYNSYSFWLASLNPFRMFRKPIPTVKAGELEKIISRCQVCM 240  
 QY 241 KKH 244  
 Db 241 KKH 244

RESULT 4  
 AAY67942  
 ID AAY67942 standard; protein; 245 AA.  
 XX AC AAY67942;  
 XX DT 03-APR-2000 (first entry)  
 XX DE Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.  
 XX KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;  
 KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;  
 KW ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasis;  
 KW myocardial angiogenesis; plaque neovascularisation; angiofibroma;  
 KW atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;  
 KW contraception; obesity.  
 XX OS Homo sapiens.  
 XX PN WO9965940-A1.  
 XX PD 23-DEC-1999.  
 XX PF 17-JUN-1999; 99WO-US013737.  
 XX PR 17-JUN-1998; 98US-0089689P.  
 XX PR 25-MAR-1999; 99US-0126175P.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX PD WPI; 2000-097708/08.  
XX DR N-PSDB; AA257158.  
XX DR Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3  
PT chain of Type IV collagen used in, e.g. treatment of benign tumors and  
PT rheumatoid arthritis.  
XX PS Claim 32; Fig 16B; 117pp; English.  
XX PS The present sequence represents the human type IV collagen alpha 3 chain.  
XX CC The present invention describes an isolated protein chosen from the NC1  
CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a  
CC fragment, analogue, derivative or mutant, which has anti-angiogenic  
CC properties. The anti-angiogenic proteins, multimers and chimeras are  
CC useful for inhibiting angiogenic activity in mammalian tissue, especially  
CC for treating diseases chosen from angiogenesis-dependent cancers, benign  
CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular  
CC angiogenesis diseases, Osler-Webber Syndrome, myocardial angiogenesis,  
CC plaque neovascularisation, telangiectasia, haemophilic joints,  
CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori  
CC ulcers, dialysis graft vascular access stenosis, contraception and  
CC obesity. The compositions can be used to inhibit a disease characterised  
CC by angiogenic activity, in conjunction with radiation therapy,  
CC chemotherapy or immunotherapy  
XX SQ Sequence 245 AA;  
Query Match 100.0%; Score 244; DB 3; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.2e-241; Mismatches 0; Indels 0; Gaps 0;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKRGDGS PATWTRGRFVTRHSQTTPAIPSCPEGTVPVLYSGFSLFVQGNRAHQD 60  
DB 2 GLKRGDGS PATWTRGRFVTRHSQTTPAIPSCPEGTVPVLYSGFSLFVQGNRAHQD 61  
QY 61 LGTGLSCLORFTHMPFLFCNVNDVNCNFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 120  
DB 62 LGTGLSCLORFTHMPFLFCNVNDVNCNFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 121  
QY 121 RCTVCEGPAIAVHSQTTPDIPCPHGWSLWKGFSLNFTSAGSEGTGQALASPGSCLE 180  
DB 122 RCTVCEGPAIAVHSQTTPDIPCPHGWSLWKGFSLNFTSAGSEGTGQALASPGSCLE 181  
QY 181 EFRASPFLECHRGTCNYYSNSYSFWLASLNPERMFRKPIPTVXAGELEKIISRCQVCM 240  
DB 182 EFRASPFLECHRGTCNYYSNSYSFWLASLNPERMFRKPIPTVXAGELEKIISRCQVCM 241  
QY 241 KKH 244  
DB 242 KKH 245  
RESULT 5  
AAU75589  
ID AAU75589 standard; protein; 245 AA.  
XX AC AAU75589;  
XX XX  
XX 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain, 'Tumstatin'.  
XX XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goopasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour.  
XX XX

OS Homo sapiens.  
XX FN WO200151523-A2.  
XX XX 19-JUL-2001.  
XX PD 08-JAN-2001; 2001WO-US0000565.  
XX PF 07-JAN-2000; 2000US-00479118.  
XX PR 04-APR-2000; 2000US-00543371.  
XX PR 21-JUL-2000; 2000US-00625191.  
XX XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PA Kalluri R;  
XX PI WPI; 2002-188037/24.  
XX DR N-PSDB; ABK15365.  
XX DR A non-Goopasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX PT Claim 29; Fig 18B; 205pp; English.  
XX PS The invention relates to a non-Goopasture fragment of alpha3(IV)NC1  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; or (b) by  
XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX endothelial cell apoptosis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; (2) use of  
XX an antibody or peptide that specifically binds the alpha1, alpha2,  
XX alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
XX preparation of a medicament for inhibiting angiogenesis or cell  
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX of a medicament for treating a proliferative disease in a vertebrate,  
XX where the disease is characterised by angiogenesis that is mediated by  
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX the presence of a medicament for promoting angiogenesis in a tissue; and  
XX (5) use of integrins in the preparation of a medicament for promoting or  
XX inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
XX or allelic variants are useful in the preparation of a medicament for  
XX treating a disorder involving inhibiting angiogenesis in a tissue, where  
XX the angiogenesis is mediated by one or more endothelial cell integrins or  
XX one or more endothelial cell integrin subunits; or by promoting or  
XX inducing endothelial cell apoptosis in a tissue, where the endothelial  
XX cell apoptosis is mediated by one or more endothelial cell integrins or  
XX one or more endothelial cell integrin subunits. The medicament is useful  
XX in inhibiting tumour growth and for the regression of an established  
XX tumour. The present sequence represents the amino acid sequence of human  
XX type IV collagen alpha 3 chain  
SQ Sequence 245 AA;

Query Match 100.0%; Score 244; DB 5; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.2e-241;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKRGDGS PATWTRGRFVTRHSQTTPAIPSCPEGTVPVLYSGFSLFVQGNRAHQD 60  
DB 2 GLKRGDGS PATWTRGRFVTRHSQTTPAIPSCPEGTVPVLYSGFSLFVQGNRAHQD 61  
QY 61 LGTGLSCLORFTHMPFLFCNVNDVNCNFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 120



Db 62 LGTGLGSLQRFRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPNMAPITGRALEPYIS 121  
QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSPFMTSAGSEGTGOALASPGSCLE 180  
Db 122 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSPFMTSAGSEGTGOALASPGSCLE 181  
QY 181 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240  
Db 182 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 241  
QY 241 KKEH 244  
Db 242 KKEH 245

## RESULT 6

ADD47063  
ID ADD47063 standard; protein; 1670 AA.

XX AC ADD47063;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein NP\_000082, SEQ ID NO 12751.

XX KW Human; pain; neuronal tissue; gene therapy;

XX KW spinal segmental nerve injury; chronic constriction injury; CCI;

XX KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; NP\_000082.

XX PT New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX PS Claim 1; Page; 1017pp; English.

XX CC The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also, claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1670 AA;

Query Match 100.0%; Score 244; DB 7; Length 1670;

Best Local Similarity 100.0%; Pred. No. 6.6e-241;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGKRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHGQD 60

Db 1427 GLKGKRGDSGSPATWTTTRGFVTRHSQTTAIPSCPGTVPVLYSGFSLFVQGNQRAHGQD 1486

QY 61 LGTGLGSLQRFRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPNMAPITGRALEPYIS 120

Db 1487 LGTGLGSLQRFRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPNMAPITGRALEPYIS 1546

QY 121 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSPFMTSAGSEGTGOALASPGSCLE 180

Db 1547 RCTVCEGPAIAIAVHSQTTDIPCPHGWISLWKGFSPFMTSAGSEGTGOALASPGSCLE 1606

QY 181 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240

Db 1607 EFRASPFLECHGRGTCNYNSYSFNLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 1666

QY 241 KKEH 244

Db 1667 KKEH 1670

## RESULT 7

ABG79218

ID ABG79218 standard; protein; 244 AA.

XX AC ABG79218;

XX DT 15-NOV-2002 (first entry)

XX DE Human type IV collagen NC1 domain mutant, alpha3(IV)NC1A1a9.

XX KW Goodpasture antigen binding protein; Goodpasture syndrome;

XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;

XX KW autoimmune condition; phosphorylation; myelin basic protein; MBP;

XX KW alpha3 type IV collagen non-collagenous domain; NC1, multiple sclerosis;

XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;

XX KW pemphigoid; lichen planus; human; mutant; mutein.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200261430-A2.

XX PD 08-AUG-2002.

XX PF 31-JAN-2002; 2002WO-EP001010.

XX PR 31-JAN-2001; 2001US-0265249P.

XX PA (SAUS/) SAUS J.

XX PI Saus J;

XX DR WPI; 2002-619280/66.

XX DR N-PSDB; ABS64503.

XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.

PT Goodpasture syndrome or lupus, comprises identifying compounds that  
PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.

XX Claim 21; Page 212-213; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) which is selected from Goodpasture antigen  
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly, or reduce formation of  
CC conformational isomers of the second target protein (II) (selected from  
CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC identical to the wild type alpha3 type IV collagen NCI domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NC1)  
CC domain (also known as the GP antigen) mutant

XX Sequence 244 AA;

Query Match 96.3%; Score 235; DB 5; Length 244;  
Best Local Similarity 100.0%; Pred. No. 2e-232;  
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHGODLGTGSCLO 69

Db 10 GSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHGODLGTGSCLO 69

QY 70 RFTTMPFLFCNVNDVNCNFAFNNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPA 129

Db 70 RFTTMPFLFCNVNDVNCNFAFNNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPA 129

QY 130 IATAVHSQTTDIPPCPHGWISLWKGSFTMFTSAGSEGTQALASPGSCLEEFPRASPFLE 189

Db 130 IATAVHSQTTDIPPCPHGWISLWKGSFTMFTSAGSEGTQALASPGSCLEEFPRASPFLE 189

QY 190 CHGRGTCNYSNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCMKKRH 244

Db 190 CHGRGTCNYSNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCMKKRH 244

RESULT 8

ABG79217

ID ABG79217 standard; protein; 244 AA.

XX ABG79217;

XX 15-NOV-2002 (first entry)

XX Human type IV collagen NCI domain mutant, alpha3(IV)NC1asp9.

XX Goodpasture antigen binding protein; Goodpasture syndrome;  
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;  
KW alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;  
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
KW pemphigoid; lichen planus; human; mutant; mutein.

XX Homo sapiens.

XX Synthetic.

XX W0200261430-A2.

XX

PD 08-AUG-2002.

XX 31-JAN-2002; 2002WO-BP001010.

XX 31-JAN-2001; 2001US-0265249P.

XX (SAUS/) SAUS J.

XX Saus J;

XX WPI; 2002-619280/66.

XX N-PSDB; ABS64502.

PT Identifying candidate compounds for treating autoimmune conditions, e.g.

PT Goodpasture syndrome or lupus, comprises identifying compounds that

PT reduce phosphorylation of, or formation of conformational isomers of,

PT target proteins.

XX Claim 21; Page 209-210; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) which is selected from Goodpasture antigen  
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly, or reduce formation of  
CC conformational isomers of the second target protein (II) (selected from  
CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC identical to the wild type alpha3 type IV collagen NCI domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NC1)  
CC domain (also known as the GP antigen) mutant

XX Sequence 244 AA;

Query Match 96.3%; Score 235; DB 5; Length 244;  
Best Local Similarity 100.0%; Pred. No. 2e-232;  
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHGODLGTGSCLO 69

Db 10 GSPATWTRGTFVTRHSQTTAIPSCPEGTVPLYSFGSFLFVQGNQRAHGODLGTGSCLO 69

QY 70 RFTTMPFLFCNVNDVNCNFAFNNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPA 129

Db 70 RFTTMPFLFCNVNDVNCNFAFNNDYSYWLSTPALMPNMNMAPITGRALEPIYSRCTVCEGPA 129

QY 130 IATAVHSQTTDIPPCPHGWISLWKGSFTMFTSAGSEGTQALASPGSCLEEFPRASPFLE 189

Db 130 IATAVHSQTTDIPPCPHGWISLWKGSFTMFTSAGSEGTQALASPGSCLEEFPRASPFLE 189

QY 190 CHGRGTCNYSNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCMKKRH 244

Db 190 CHGRGTCNYSNSYSYFWLASLNPFRMFRKPIPTSTVKAGELEKIISRCQVCMKKRH 244

RESULT 9

AAU75596

ID AAU75596 standard; protein; 191 AA.

XX AAU75596;

XX 08-MAY-2002 (first entry)

XX

DE Human type IV collagen alpha 3 chain mutant, Tumstatin N53.  
XX Human; type IV collagen alpha 3 chain; cytosolic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX Tumstatin; angiogenesis; tumour; mutin; mutant.  
XX Homo sapiens.  
XX WO200151523-A2.  
XX 19-JUL-2001.  
XX 08-JAN-2001; 2001WO-US000565.  
XX 07-JAN-2000; 2000US-00479118.  
XX 04-APR-2000; 2000US-00543371.  
XX 21-JUL-2000; 2000US-00625191.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.  
XX Example 32; Page; 205pp; English.  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; or (b) by  
XX promoting or inducing endothelial cell apoptosis in a tissue, where the  
XX endothelial cell apoptosis is mediated by one or more endothelial cell  
XX integrins or one or more endothelial cell integrin subunits; (2) use of  
XX an antibody or peptide that specifically binds the alpha1, alpha2,  
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
XX preparation of a medicament for inhibiting angiogenesis or cell  
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody  
XX fragment or peptide of receptor-mediated angiogenesis in the preparation  
XX of a medicament for treating a proliferative disease in a vertebrate,  
XX where the disease is characterised by angiogenesis that is mediated by  
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors  
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
XX the presence of a medicament for promoting angiogenesis in a tissue; and  
XX (5) use of integrins in the preparation of a medicament for promoting or  
XX inducing angiogenesis or cell proliferation in a tissue. The fragments  
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
XX or allelic variants are useful in the preparation of a medicament for  
XX treating a disorder involving inhibiting angiogenesis in a tissue, where  
XX the angiogenesis is mediated by one or more endothelial cell integrins or  
XX one or more endothelial cell integrin subunits; or by promoting or  
XX inducing endothelial cell apoptosis in a tissue, where the endothelial  
XX cell apoptosis is mediated by one or more endothelial cell integrins or  
XX one or more endothelial cell integrin subunits. The medicament is useful  
XX in inhibiting tumour growth and for the regression of an established  
XX tumour. The present sequence represents the amino acid sequence of human  
XX type IV collagen alpha 3 chain mutant, Tumstatin N53, which consists of  
XX residues 54-244 of Tumstatin. Note: The present sequence is not shown in  
XX the specification but is derived from the wild type human Tumstatin  
XX sequence given in figure 18A (see AAU75589)  
XX Sequence 191 AA;

Query Match 78.3%; Score 191; DB 5; Length 191;  
Best Local Similarity 100.0%; Pred. No. 2.3e-187;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 53 NQRAHGQDLTGSLCQRFTTTPFLECNVNDVNCNPFASRNDYSYWLSTPALMPNVPATG 112  
DB 1 NQRAHGQDLTGSLCQRFTTTPFLECNVNDVNCNPFASRNDYSYWLSTPALMPNVPATG 60  
QY 113 RALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGEGTGQAL 172  
DB 61 RALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGEGTGQAL 120  
QY 173 ASFGSCLEBERASPFLECHGRTGTCNYSNSYSFSLASLNPFRMFRKPISTVKAGLEKI 232  
DB 121 ASFGSCLEBERASPFLECHGRTGTCNYSNSYSFSLASLNPFRMFRKPISTVKAGLEKI 180  
QY 233 ISRCQVCMKKR 243  
DB 181 ISRCQVCMKKR 191  
RESULT 10  
ADA20260  
ID ADA20260 standard; protein; 191 AA.  
XX AC ADA20260;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tum-1 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tum-1;  
XX tumstatin N53.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 22; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumor growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which  
XX specifically bind anti-angiogenic proteins and peptides (in particular  
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX collagen) are disclosed. The proteins of the invention may inhibit tumour

CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequence of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-1 (tumstatin N53), an abridged form of the  
 CC "tumstatin" protein of the invention which was derived from the amino  
 CC acid sequence of the alpha 3 chain of human type IV collagen. Note: This  
 CC sequence (Seq ID22) does not appear in the specification but was created  
 CC by the indexer from information given in the specification.  
 XX  
 XX Sequence 191 AA;

Query Match 78.3%; Score 191; DB 6; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-187;  
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 54 QRAHGDLGLGSLQRTFTMPFLFCNVNDVNCNFSRNDYSVWLSPTALPMNVPITGR 113  
 DB 1 QRAHGDLGLGSLQRTFTMPFLFCNVNDVNCNFSRNDYSVWLSPTALPMNVPITGR 60  
 QY 114 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSTFMFTSAGSEGTGQALA 173  
 DB 61 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSTFMFTSAGSEGTGQALA 120  
 QY 174 SPGSCLEBEPFRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVKAGELEKII 233  
 DB 121 SPGSCLEBEPFRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVKAGELEKII 180  
 QY 234 SRCQVCMKKRH 244  
 DB 181 SRCQVCMKKRH 191

RESULT 11  
 AAU75598  
 ID AAU75598 standard; protein; 254 AA.

XX AAU75598;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 3 chain mutant, Tum-3.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.

PN WO200151523-A2.

PD 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.

XX Example 36; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tum-3, which consists of residues  
 CC 133-244 of Tumstatin. Note: The present sequence is not shown in the  
 CC specification but is derived from the wild type human Tumstatin sequence  
 CC given in figure 18A (see AAU75598)

XX Sequence 254 AA;

Query Match 75.0%; Score 183; DB 5; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-179;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LGTGLSCLQRTFTMPFLFCNVNDVNCNFSRNDYSVWLSPTALPMNVPITGRALPEYIS 120

DB 72 LGTGLSCLQRTFTMPFLFCNVNDVNCNFSRNDYSVWLSPTALPMNVPITGRALPEYIS 131

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSTFMFTSAGSEGTGQALASPGSCLE 180

DB 132 RCTVCEGPAIAIAVHSQTTDIPPCPHGWSLWKGFSTFMFTSAGSEGTGQALASPGSCLE 191

QY 181 EPRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVKAGELEKIIISRCQVCM 240

DB 192 EPRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPTVKAGELEKIIISRCQVCM 251

QY 241 KKR 243

DB 252 KKR 254

RESULT 12

AAU75598

ID AAY31993 standard; protein; 268 AA.

XX AAY31993;

XX 05-JAN-2000 (first entry)

XX Type IV collagen NC1 domain alpha-3 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;

KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;

KW rheumatoid arthritis; retinal neovascularization;  
 KW choroidal neovascularization; macular degeneration;  
 KW corneal neovascularization; retinopathy of prematurity;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidermal keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; sogrens; acne rosacea; phlyctenulosis;  
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;  
 KW ulcer; herpes simplex infection; Herpes zoster infection;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW Steven's Johnson disease; Wegener's sarcoidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;  
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Bales disease; Bechets disease;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;  
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;  
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;  
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;  
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;  
 KW pemphigoid.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX  
 PH Key Location/Qualifiers  
 FT Peptide 1..17 /note= "BM40 signal peptide"  
 FT Protein 18..268 /note= "mature protein"  
 FT Peptide 18..25 /note= "affinity tag"  
 FT Protein 26..268 /note= "NC1 alpha-3 monomer"  
 FT  
 XX WO9949885-A2.  
 XX  
 XX 07-OCT-1999.  
 XX  
 XX 26-MAR-1999; 99WO-US006445.  
 XX  
 XX 27-MAR-1998; 98US-0079783P.  
 XX 29-OCT-1998; 98US-0106170P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 XX  
 XX Hudson BG, Sarraz MP;  
 XX  
 XX WPI; 1999-601297/51.  
 XX N-PSDB; AA220091.  
 XX  
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer  
 XX useful for treating e.g. tumor growth or metastasis, neovascularisation,  
 XX etc.  
 XX  
 XX Disclosure; Fig 17c; 56pp; English.  
 XX  
 XX This sequence represents a recombinant type IV collagen non-collagenous  
 XX (NC1) domain alpha-3 polypeptide composed of a BM40 signal sequence  
 XX (which is cleaved from the mature protein) to facilitate protein  
 XX secretion, and a mature protein comprising an affinity tag (facilitates  
 XX purification and identification of the material) and the alpha-1 chain  
 XX monomer. The invention provides methods and kits for inhibiting  
 XX angiogenesis, tumor growth and metastasis, and endothelial cell  
 XX interaction with the extracellular matrix, each method comprising  
 XX contacting the tumor or animal tissue with 1 or more isolated type IV  
 XX collagen NC1 alpha chain monomer(s) selected from the group consisting of  
 XX alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-  
 XX 96). The monomers can be produced via recombinant protein expression. The  
 XX polynucleotides and polypeptides are used to treat an angiogenesis-  
 XX mediated disorder or condition, especially selected from solid and blood-  
 XX borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal

CC neovascularization, choroidal neovascularization, macular degeneration,  
 CC corneal neovascularization, retinopathy of prematurity, corneal graft  
 CC rejection, neovascular glaucoma, retrolental fibroplasia, epidermal  
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic  
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, sogrens,  
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid  
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes  
 CC simplex infections, herpes zoster infections, protozoan infections,  
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal  
 CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's  
 CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,  
 CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,  
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic  
 CC uveitis, chronic vitritis, Lyme's disease, Bales disease, Bechets  
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic  
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser  
 CC complications, abnormal proliferation of fibrovascular tissue,  
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,  
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative  
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)  
 XX  
 SQ Sequence 268 AA;  
 Query Match 66.8%; Score 163; DB 2; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-158;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 5 KRGDGSPATWTTTRGFVTRHSQTTAIPSCPEGTPLYSGFSLFVQGNQRAHQDGLTL 64  
 Db 29 KRGDGSPATWTTTRGFVTRHSQTTAIPSCPEGTPLYSGFSLFVQGNQRAHQDGLTL 88  
 Qy 65 GSCLOQRTTTPFLLFCNVNDVCNPFASRNDYSYMLSTPALMPMNPAPITGRALPEYISRCTV 124  
 Db 89 GSCLOQRTTTPFLLFCNVNDVCNPFASRNDYSYMLSTPALMPMNPAPITGRALPEYISRCTV 148  
 Qy 125 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 167  
 Db 149 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 191  
 RESULT 13  
 AAY97555  
 ID AAY97555 standard; protein; 268 AA.  
 XX  
 AC AAY97555;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 XX Human alpha3(IV)NC1 protein sequence.  
 DE  
 XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;  
 KW tumor growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;  
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;  
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;  
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;  
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;  
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;  
 KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200005932-A1.  
 FN  
 XX 12-OCT-2000.  
 PD  
 XX 31-MAR-2000; 2000WO-US008678.  
 PF  
 XX 01-APR-1999; 99US-0127391P.  
 FR  
 XX (BIOS-) BIOSTRATUM INC.  
 FA  
 XX Brooks P, Hudson B;  
 PI  
 XX





DR N-PSDB; ABS64491.  
XX  
PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
PT Goodpasture syndrome or lupus, comprises identifying compounds that  
PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.  
XX  
PS Example 3; Page 199-200; 217pp; English.  
XX  
CC The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) (which is selected from Goodpasture antigen  
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCL)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of  
CC conformational isomers of the second target protein (II) (selected from  
CC an alpha3 type IV collagen NCL domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC NCL domain conformational isomer, which has an amino acid sequence  
CC identical to the wild type alpha3 type IV collagen NCL domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated  
CC type IV collagen alpha3 NCL domain. The human gene for GPBP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NCL)  
CC domain (also known as the GP antigen) or an MBP isoform  
XX  
SQ Sequence 211 AA;  
Query Match 65.2%; Score 159; DB 5; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.7e-154;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQD 60  
DB 1 GLKGRGSGSPATWTRGFTVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQD 60  
QY 61 LGTLGSLQRTFTMPLFCNVNDVNCNPNFASNDYSYWLSTPALMPMNPATITGRALEPIYS 120  
DB 61 LGTLGSLQRTFTMPLFCNVNDVNCNPNFASNDYSYWLSTPALMPMNPATITGRALEPIYS 120  
QY 121 RCTVCEGPAIAVAHSQTDDIPPCPHGWISLWKGFSTFM 159  
DB 121 RCTVCEGPAIAVAHSQTDDIPPCPHGWISLWKGFSTFM 159  
RESULT 16  
ADCL7697  
ID ADCL7697 standard; protein; 232 AA.  
XX  
AC ADCL7697;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human type IV collagen alpha 3 chain protein SEQ ID NO:304.  
XX  
KW crystallised NCL domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antanaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
OS Homo sapiens.  
XX

PN WO2003012122-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 26-JUL-2002; 2002WO-US023763.  
XX  
PR 27-JUL-2001; 2001US-0308523P.  
PR 29-OCT-2001; 2001US-0351289P.  
PR 22-MAR-2002; 2002US-0366854P.  
PR 03-JUN-2002; 2002US-0385362P.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
PA (SUND/) SUNDARAMOORTHY M.  
PA (HUDS/) HUDSON B.  
PI Sundaramoorthy M, Hudson B;  
DR WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.  
XX  
PS Disclosure; SEQ ID NO 304; 168pp; English.  
XX  
CC The present invention describes a crystallised NCL domain hexamer of type  
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NCL  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NCL domain hexamer of type IV collagen (I) has cytostatic,  
CC antipsoriatic, antanaemic, ophthalmological, antiarteriosclerotic and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis in  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents an amino acid sequence which is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 232 AA;  
Query Match 63.5%; Score 155; DB 7; Length 232;  
Best Local Similarity 100.0%; Pred. No. 2.3e-150;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 ATWTRGFTVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQDGLTGLSCLOPFT 72  
DB 1 ATWTRGFTVTRHSQTATPSCPEGTVPLYSFSLFVQGNRAHQDGLTGLSCLOPFT 60  
QY 73 TWPFLECNVNDVNCNPNFASNDYSYWLSTPALMPMNPATITGRALEPIYSRCTVCEGPAIAI 132  
DB 61 TWPFLECNVNDVNCNPNFASNDYSYWLSTPALMPMNPATITGRALEPIYSRCTVCEGPAIAI 120  
QY 133 AVHSQTDDIPPCPHGWISLWKGFSTFMFTSAGSEG 167  
DB 121 AVHSQTDDIPPCPHGWISLWKGFSTFMFTSAGSEG 155  
RESULT 17  
AAAY44172  
ID AAAY44172 standard; protein; 218 AA.



```

XX AC AAY44172;
XX DT 01-FEB-2000 (first entry)
XX DE Human type IV collagen alpha3 chain protein.
XX KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;
XX KW Goodpasture syndrome; antibody; Blood; tissue; human; nephrotrophism.
XX OS Homo sapiens.
XX PN US5973120-A.
XX PD 26-OCT-1999.
XX PF 07-MAR-1995; 95US-00399889.
XX PR 30-NOV-1990; 90US-00621091.
XX PA (UYVA ) UNIV YALE.
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX PI Hudson BG, Reenders ST, Morrison KE;
XX WI; 1999-610317/52.
XX DR N-PSDB; AAZ28775.
XX PT Isolated alpha 3 chain of type IV collagen polypeptide useful for
XX PT diagnosis and treatment of Goodpasture syndrome.
XX PS Claim 2; Col 35-36; 27pp; English.
XX CC This sequence represents a recombinant human alpha3 chain of type IV
XX CC collagen polypeptide. The sequence corresponds to the 218 amino acids of
XX CC the C-terminal non-collagenous domain. Alpha3 chain collagen polypeptides
XX CC are useful for detecting Goodpasture antibodies in blood or tissue from a
XX CC human patient and for treating Goodpasture syndrome, especially by
XX CC neutralising the antibodies in the blood. The polypeptides also have a
XX CC nephrotrophic activity
XX SQ Sequence 218 AA;
Query Match 57.8%; Score 141; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 QTTAIPSCPGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLQRTTTPFLFCNVNDVCN 86
DB 1 QTTAIPSCPGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLQRTTTPFLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEG 167
DB 121 GWISLWKGFSPIMFTSAGSEG 141
RESULT 18
AAY56784
ID AAY56784 standard; protein; 218 AA.
XX AC AAY56784;
XX DT 27-MAR-2000 (first entry)
XX DE Human alpha3 type IV collagen C-terminal domain.
XX KW Goodpasture syndrome; type IV collagen; alpha3 chain; human.
XX OS Homo sapiens.

```

```

XX PN US6007980-A.
XX PD 28-DEC-1999.
XX PF 07-OCT-1998; 98US-00167364.
XX PR 30-NOV-1990; 90US-00621091.
XX PR 07-MAR-1995; 95US-00399889.
XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.
XX PA (UYVA ) UNIV YALE.
XX PI Hudson BG, Reenders ST, Morrison KE;
XX WI; 2000-096371/08.
XX DR N-PSDB; AAZ46729.
XX PT Diagnosing and treating Goodpasture syndrome using a peptide derived from
XX PT type IV collagen.
XX PS Disclosure; Col 23-26; 26pp; English.
XX CC The invention provides a method of detecting Goodpasture antibodies in
XX CC the fluid of a patient by contacting it with a peptide comprising at most
XX CC 218 amino acids of the human alpha3 chain type IV collagen that contains
XX CC the fragment shown in AAY56785. The methods are useful for the diagnosis
XX CC and treatment of Goodpasture syndrome. The present sequence represents
XX CC the carboxy terminal noncollagenous domain of the human alpha3 chain of
XX CC type IV collagen
XX SQ Sequence 218 AA;
Query Match 57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.1e-136;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 QTTAIPSCPGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLQRTTTPFLFCNVNDVCN 86
DB 1 QTTAIPSCPGTVPVLYSGFSFLVQGNQRAHQDLGLTGLSCLQRTTTPFLFCNVNDVCN 60
QY 87 FASRNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALMPMNPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120
QY 147 GWISLWKGFSPIMFTSAGSEG 167
DB 121 GWISLWKGFSPIMFTSAGSEG 141
RESULT 19
AAE09484
ID AAE09484 standard; protein; 218 AA.
XX AC AAE09484;
XX DT 19-NOV-2001 (first entry)
XX DE Human alpha-3 chain of type IV collagen protein.
XX KW Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
XX KW Goodpasture syndrome.
XX OS Homo sapiens.
XX PN US6277558-B1.
XX PD 21-AUG-2001.
XX PF 12-NOV-1999; 99US-00439897.
XX PR 30-NOV-1990; 90US-00621091.
XX PR 07-MAR-1995; 95US-00399889.

```



PR 07-JAN-2000; 2000US-00479118.  
PR 04-APR-2000; 2000US-00543371.  
PR 21-JUL-2000; 2000US-00625191.  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
DR A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX  
XX  
XX Claim 31; Page 152; 205pp; English.  
XX  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterized by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
XX  
XX Sequence 132 AA;  
XX  
XX Query Match 53.7%; Score 131; DB 5; Length 132;  
XX Best Local Similarity 100.0%; Pred. No. 5.9e-126;  
XX Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKKGKDGSGPATWTRGTFVTRHSQTTAIPSCPEGTVPLVSGSFVQGNQAHQGD 60  
Db 2 GLKKGKDGSGPATWTRGTFVTRHSQTTAIPSCPEGTVPLVSGSFVQGNQAHQGD 61  
QY 61 LGTIGSLQRFMTMPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALEPIS 120  
Db 62 LGTIGSLQRFMTMPFLFCNVNDVNCNFASRNDYSYWLSTPALMPMNPAPITGRALEPIS 121  
QY 121 RCTVCBGAIPA 131  
Db 122 RCTVCBGAIPA 132

RESULT 22  
AAU75594  
ID AAU75594 standard; protein; 124 AA.  
XX AC AAU75594;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human type IV collagen alpha 3 chain mutant, Tumstatin 333.  
XX KW Human; type IV collagen alpha 3 chain; cystostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutant; mutant.  
XX OS Homo sapiens.  
XX PN WO200151523-A2.  
XX PD 19-JUL-2001.  
XX PF 08-JAN-2001; 2001WO-US000565.  
XX PR 07-JAN-2000; 2000US-00479118.  
PR 04-APR-2000; 2000US-00543371.  
PR 21-JUL-2000; 2000US-00625191.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2002-188037/24.  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.  
XX Example 33; Page: 205pp; English.  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterized by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
XX

CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tumstatin 333, which consists of  
CC residues 2-125 of Tumstatin. Note: The present sequence is not shown in  
CC the specification but is derived from the wild type human Tumstatin  
CC sequence given in figure 18A (see AAU75589)

XX  
SQ Sequence 124 AA;

Query Match 50.8%; Score 124; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 8.5e-119;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60  
Dy |||||  
1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60  
QY 61 LGTLGSLQRFVTFMPLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120  
Dy |||||  
61 LGTLGSLQRFVTFMPLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120  
QY 121 RCTV 124  
Dy |||||  
121 RCTV 124

## RESULT 23.

ADA20258.  
ID ADA20258 standard; protein; 124 AA.

XX AC ADA20258;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tumstatin 333 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 333.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX XX 24-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;

XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.  
XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 20; 240pp; English.

XX CC This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX CC Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2

CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tumstatin 333, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
CC ID20) does not appear in the specification but was created by the indexer  
CC from information given in the specification.

XX SQ Sequence 124 AA;

Query Match 50.8%; Score 124; DB 6; Length 124;  
Best Local Similarity 100.0%; Pred. No. 8.5e-119;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60  
Dy |||||  
1 GLKGRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGLFVQGNQRAHQD 60  
QY 61 LGTLGSLQRFVTFMPLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120  
Dy |||||  
61 LGTLGSLQRFVTFMPLFCNVNDVCFASRNDYSYWLSTPALMPMNMAPIITGRALEPYIS 120  
QY 121 RCTV 124  
Dy |||||  
121 RCTV 124

## RESULT 24

ADA20259.  
ID ADA20259 standard; protein; 120 AA.

XX AC ADA20259;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human tumstatin deletion protein tumstatin 334 amino acid sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytosolic; gene therapy; alpha 3 chain; tumstatin; human; tumstatin 334.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX XX 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;

XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20224.

XX PT New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 94; SEQ ID NO 21; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA

CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequence of the invention may be useful in gene therapy. The present  
 CC sequence is that of tumstatin 334, an abridged form of the "tumstatin"  
 CC protein of the invention which was derived from the amino acid sequence  
 CC of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq  
 CC ID21) does not appear in the specification but was created by the indexer  
 CC from information given in the specification.

XX Sequence 120 AA;

Query Match 49.2%; Score 120; DB 6; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-114;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 184

Db 1 CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 60

QY 185 SPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCMKKRH 244

Db 61 SPFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCMKKRH 120

RESULT 25

ID ADA20262  
 ADADA20262 standard; protein; 112 AA.

XX ADA20262;

AC ADA20262;

DT 20-NOV-2003 (first entry)

XX Human tumstatin deletion protein tum-3 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-3.

XX Homo sapiens.

OS WO2003059257-A2.

FN 24-JUL-2003.

PD 20-DEC-2002; 2002WO-US040938.

PF 21-DEC-2001; 2001US-00032221.

PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA Kalluri R;

PI WPI; 2003-587256/55.

DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor

PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Example 35; SEQ ID NO 24; 240pp; English.

XX

CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequence of the invention may be useful in gene therapy. The present  
 CC sequence is that of tum-3, an abridged form of the "tumstatin" protein of  
 CC the invention which was derived from the amino acid sequence of the alpha  
 CC 3 chain of human type IV collagen. Note: This sequence (Seq ID24) does  
 CC not appear in the specification but was created by the indexer from  
 CC information given in the specification.

XX Sequence 112 AA;

Query Match 45.9%; Score 112; DB 6; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-106;  
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 192

Db 1 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGQALASPGSCLEEFRA 60

QY 193 RGTCTNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCMKKRH 244

Db 61 RGTCTNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKIISRCQVCMKKRH 112

RESULT 26

AAAR79164  
 ID AAR79164 standard; protein; 218 AA.

XX AAR79164;

AC AAR79164;

DT 22-DEC-1995 (first entry)

XX Partial sequence of human alpha 3 chain of type IV collagen.

DE Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.

OS Homo sapiens.

XX US5424408-A.

PN 13-JUN-1995.

XX 30-NOV-1990; 90US-00621091.

PF 30-NOV-1990; 90US-00621091.

PR (UYVA ) UNIV YALE.

PA (UNIV ) UNIV KANSAS MEDICAL CENT.

PI Morrison KE, Reeders ST, Hudson BG;

XX WPI; 1995-262631/34.

DR N-PSDB; AAQ96291.

XX cDNA's encoding human or bovine alpha-3 type 4 collagen peptide(s) -  
 PT useful for detection and therapeutic removal of antibodies associated  
 PT with Goodpasture syndrome.

XX Disclosure; Col 7-10; 33pp; English.

XX

CC Using the PCR with primers derived from each end of the known 27 AA  
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp bovine  
 CC genomic fragment was amplified. This fragment was then used to a bovine  
 CC lens cDNA library and a 1.5 kb partial cDNA clone was obtained. (clone  
 CC KMC15). This encodes 238 residues of the triple helical collagenous  
 CC domain and all 233 residues of the C-terminal non-collagenous (NC1)  
 CC domain of the alpha 3 (IV) chain. This bovine cDNA clone was used to  
 CC screen a human kidney cDNA library and a 2.7 kb human cDNA clone (clone  
 CC KMC27) was obtained. This clone encodes 218 residues of the NC1 domain and a  
 CC portion of the 3' UTR region of the human alpha 3 (IV) chain. The COL4A3  
 CC gene localises to chromosome 2 and therefore mutations in COL4A3 cannot  
 CC be responsible for Alport syndrome which is X-linked. An isolated and  
 CC substantially pure nt. having the sequence in A096291 is claimed  
 XX  
 XX Sequence 218 AA;  
 CC  
 CC Query Match 40.6%; Score 99; DB 2; Length 218;  
 CC Best Local Similarity 100.0%; Pred. No. 6.4e-93;  
 CC Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 27 QTTAIPSCPGTPLYSGFSLFVQGNQRAHGQDLGLGSCLORFMTTLPFCNVNDVCN 86  
 CC Db 1 QTTAIPSCPGTPLYSGFSLFVQGNQRAHGQDLGLGSCLORFMTTLPFCNVNDVCN 60  
 CC  
 CC QY 87 FASNDYSYMLSTPALPMNMAPITGTRALEPYISRCTVC 125  
 CC Db 61 FASNDYSYMLSTPALPMNMAPITGTRALEPYISRCTVC 99  
 CC  
 CC RESULT 27  
 CC AAU75607  
 CC ID AAU75607 standard; protein; 88 AA.  
 CC AC AAU75607;  
 CC DT 08-MAY-2002 (first entry)  
 CC DE Human type IV collagen alpha 3 chain mutant, Tumstatin-45-132.  
 CC KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 CC KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 CC KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 CC KW Tumstatin; angiogenesis; tumour; mutin; mutant.  
 CC OS Homo sapiens.  
 CC XX WO200151523-A2.  
 CC XX 19-JUL-2001.  
 CC XX 08-JAN-2001; 2001WO-0000565.  
 CC XX 07-JAN-2000; 2000US-00479118.  
 CC XX 04-APR-2000; 2000US-00543371.  
 CC XX 21-JUL-2000; 2000US-00625191.  
 CC XX (BETH-) BETH-ISRAEL DEACONESS MEDICAL CENT.  
 CC XX Kalluri R;  
 CC XX WPI; 2002-188037/24.  
 CC DR A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 CC PT treating disorders involving angiogenesis.  
 CC PS Claim 32; Page 152; 205pp; English.  
 CC CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,

CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, Tumstatin-45-132, which consists  
 CC of residues 45-132 of Tumstatin  
 XX  
 XX Sequence 88 AA;  
 CC  
 CC Query Match 36.1%; Score 88; DB 5; Length 88;  
 CC Best Local Similarity 100.0%; Pred. No. 5.4e-82;  
 CC Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 44 GFSFLFVQGNQRAHGQDLGLGSCLORFMTTLPFCNVNDVCNPFASNDYSYMLSTPALM 103  
 CC Db 1 GFSFLFVQGNQRAHGQDLGLGSCLORFMTTLPFCNVNDVCNPFASNDYSYMLSTPALM 60  
 CC  
 CC QY 104 PNMAPITGTRALEPYISRCTVCCEGPAIA 131  
 CC Db 61 PNMAPITGTRALEPYISRCTVCCEGPAIA 88  
 CC  
 CC RESULT 28  
 CC ADA20271  
 CC ID ADA20271 standard; protein; 88 AA.  
 CC AC ADA20271;  
 CC XX 20-NOV-2003 (first entry)  
 CC DE Human tumstatin deletion protein tumstatin 45-132 amino acid sequence.  
 CC KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 CC KW metastasis; basement membrane organisation; type IV collagen network;  
 CC KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 CC KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 CC KW cytostatic; gene therapy; alpha 3 chain; tumstatin; human;  
 CC KW tumstatin 45-132.  
 CC OS Homo sapiens.  
 CC XX WO2003059257-A2.  
 CC XX 24-JUL-2003.  
 CC XX 20-DEC-2002; 2002WO-US040938.

```
PR 21-DEC-2001; 2001US-00032221.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX NPI; 2003-587256/55.
XX N-PSDB; ADA20224.
XX New peptide, useful for preparing a composition for inhibiting tumor
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX Claim 94; SEQ ID NO 33; 240pp; English.
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tumstatin 45-132, an abridged form of the "tumstatin"
XX protein of the invention which was derived from the amino acid sequence
XX of the alpha 3 chain of human type IV collagen. Note: This sequence (Seq
XX ID33) does not appear in the specification but was created by the indexer
XX from information given in the specification.
XX SQ Sequence 88 AA;
Query Match 36.1%; Score 88; DB 6; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.4e-82;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 FSLFVQGNORAHGQDLGTGSCLOQRTTTPFLFCNVNDVNCNPNASNDYSYMLSTPALMP 104.
DB 1 FSLFVQGNORAHGQDLGTGSCLOQRTTTPFLFCNVNDVNCNPNASNDYSYMLSTPALMP 60
QY 105 MNWAPITGRALEPYISRCTVCEGPATAI 132
DB 61 MNWAPITGRALEPYISRCTVCEGPATAI 88
RESULT 29
AAU75608
ID AAU75608 standard; protein; 88 AA.
AC AAU75608;
XX 08-MAY-2002 (first entry)
XX Human type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A.
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.
XX Homo sapiens.
XX Key Location/Qualifiers
XX FT Misc-difference 82
XX FT /note= "Wild type Cys substituted with Ala"
XX XX
XX W02001:51523-A2.
```

```
XX 19-JUL-2001.
XX 08-JAN-2001; 2001WO-US000565.
XX 07-JAN-2000; 2000US-00479118.
XX 04-APR-2000; 2000US-00543371.
XX 21-JUL-2000; 2000US-00625191.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX Kalluri R;
XX NPI; 2002-188037/24.
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX treating disorders involving angiogenesis.
XX Claim 41; Page 153; 205pp; English.
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alphav, betav or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
XX Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX or allelic variants are useful in the preparation of a medicament for
XX treating a disorder involving inhibiting angiogenesis in a tissue, where
XX the angiogenesis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits; or by promoting or
XX inducing endothelial cell apoptosis in a tissue, where the endothelial
XX cell apoptosis is mediated by one or more endothelial cell integrins or
XX one or more endothelial cell integrin subunits. The medicament is useful
XX in inhibiting tumour growth and for the regression of an established
XX tumour. The present sequence represents the amino acid sequence of human
XX type IV collagen alpha 3 chain mutant, Tumstatin-5-126-C-A, which
XX consists of residues 5-126 of Tumstatin
XX SQ Sequence 88 AA;
Query Match 33.2%; Score 81; DB 5; Length 88;
Best Local Similarity 100.0%; Pred. No. 8.3e-75;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 GFSFLFVQGNORAHGQDLGTGSCLOQRTTTPFLFCNVNDVNCNPNASNDYSYMLSTPALM 103
DB 1 GFSFLFVQGNORAHGQDLGTGSCLOQRTTTPFLFCNVNDVNCNPNASNDYSYMLSTPALM 60
QY 104 PMNMAPITGRALEPYISRCTV 124
DB 61 PMNMAPITGRALEPYISRCTV 81
```



```
RESULT 30
ADA20272 ID ADA20272 standard; protein; 88 AA.
AC AA
XX ADA20272;
XX
XX 20-NOV-2003 (first entry)
DT
DE Human tumstatin deletion protein tumstatin 5-125-C-A amino acid sequence.
XX
XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;
XX metastasis; basement membrane organisation; type IV collagen network;
XX C-terminal globular non-collagenous domain; NCI; type IV collagen;
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;
XX cytoskeletal; gene therapy; alpha 3 chain; tumstatin; human;
XX tumstatin 5-125-C-A; mutant; mutein.
XX
XX Synthetic.
OS
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH
FT Misc-difference 81
FT /note= "Wild-type Cys substituted by Ala at position 125
FT of full-length tumstatin"
XX
XX WO2003059257-A2.
PN
XX
XX 24-JUL-2003.
PD
XX
XX 20-DEC-2002; 2002WO-US040938.
XX
XX 21-DEC-2001; 2001US-00032221.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2003-587256/55.
XX
XX New peptide, useful for preparing a composition for inhibiting tumor
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.
XX
XX Claim 94; SEQ ID NO 34; 240pp; English.
XX
XX This invention relates to novel isolated proteins and their fragments
XX with anti-angiogenic properties. The invention also relates to the DNA
XX sequences which encode the novel proteins. A wide variety of diseases are
XX the result of undesirable angiogenesis. The formation of new capillaries
XX from pre-existing vessels is essential for tumour growth and metastasis.
XX Basement membrane organisation is dependent on the assembly of a type IV
XX collagen network which may occur through the C-terminal globular non-
XX collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2
XX forms are ubiquitously exhibited in human basement membranes. In the
XX present invention, cell surface receptors (in particular integrins) which
XX specifically bind anti-angiogenic proteins and peptides (in particular
XX the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV
XX collagen) are disclosed. The proteins of the invention may inhibit tumour
XX growth, angiogenic activity in mammalian tissue or protein synthesis in
XX endothelial cells and thus may exhibit cytostatic activity. The DNA
XX sequences of the invention may be useful in gene therapy. The present
XX sequence is that of tumstatin 5-125-C-A, a mutated and abridged form of
XX the "tumstatin" protein of the invention which was derived from the amino
XX acid sequence of the alpha 3 chain of human type IV collagen. Note: This
XX sequence (Seq ID33) does not appear in the specification but was created
XX by the indexer from information given in the specification.
XX
XX Sequence 88 AA;
SQ
Query Match 32.8%; Score 80; DB 6; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.8e-74;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 FSFLFVQGNORAHGQDLGTLGSCLOQRTTTPFFLCNNVDCNPNFASRNDYSYWLSTPALMP 104
DB 1 FSFLFVQGNORAHGQDLGTLGSCLOQRTTTPFFLCNNVDCNPNFASRNDYSYWLSTPALMP 60
OY 105 MNWAPITGRALEPVIISRCTV 124
DB 61 MNWAPITGRALEPVIISRCTV 80

RESULT 31
AAU75600
ID AAU75600 standard; protein; 79 AA.
XX
AC AAU75600;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human type IV collagen alpha 3 chain mutant, Tum-5.
XX
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;
XX non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;
XX Tumstatin; angiogenesis; tumour; mutein; mutant.
XX
XX Homo sapiens.
XX
XX WO200151523-A2.
PN
XX
XX 19-JUL-2001.
PD
XX
XX 08-JAN-2001; 2001WO-US000565.
XX
XX 07-JAN-2000; 2000US-00479118.
XX
XX 04-APR-2000; 2000US-00543371.
XX
XX 21-JUL-2000; 2000US-00625191.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI; 2002-189037/24.
XX
XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and
XX treating disorders involving angiogenesis.
XX
XX Example 40; Page: 205pp; English.
XX
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI
XX domain, having one or more of the characteristics selected from: (a) the
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX proliferation of endothelial cells; and (c) the ability to cause
XX apoptosis of endothelial cells. Also described are the following: (1) use
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX analogue or allelic variant in the preparation of a medicament for
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX where the angiogenesis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; or (b) by
XX promoting or inducing endothelial cell apoptosis in a tissue, where the
XX endothelial cell apoptosis is mediated by one or more endothelial cell
XX integrins or one or more endothelial cell integrin subunits; (2) use of
XX an antibody or peptide that specifically binds the alpha1, alpha2,
XX alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX preparation of a medicament for inhibiting angiogenesis or cell
XX proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX fragment or peptide of receptor-mediated angiogenesis in the preparation
XX of a medicament for treating a proliferative disease in a vertebrate,
XX where the disease is characterised by angiogenesis that is mediated by
XX receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX the presence of a medicament for promoting angiogenesis in a tissue; and
XX (5) use of integrins in the preparation of a medicament for promoting or
XX inducing angiogenesis or cell proliferation in a tissue. The fragments
```

CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tum-5, which consists of residues  
CC 54-132 of Tumstatin. Note: The present sequence is not shown in the  
CC specification but is derived from the wild type human Tumstatin sequence  
CC given in figure 18A (see AAU75599)  
XX  
SQ Sequence 79 AA;

Query Match 32.4%; Score 79; DB 5; Length 79;  
Best Local Similarity 100.0%; Pred. No. 8.5e-73;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 NQAHGQDLGTLGSCLOQFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITG 112  
Db 1 NQAHGQDLGTLGSCLOQFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITG 60

QY 113 RALEPYISRCTVCEGPAIA 131  
Db 61 RALEPYISRCTVCEGPAIA 79

RESULT 32  
ADA20264  
ID ADA20264 standard; protein; 79 AA.

XX ADA20264;

AC ADA20264;

DT 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tum-5 amino acid sequence.

KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cyostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-5.

XX Homo sapiens.

OS Homo sapiens.

PN WO200305257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX Claim 94; SEQ ID NO 26; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.

CC Basement membrane organisation is dependent on the assembly of a type IV,  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cyostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tum-5, an abridged form of the "tumstatin" protein of  
CC the invention which was derived from the amino acid sequence of the alpha  
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID26) does  
CC not appear in the specification but was created by the indexer from  
CC information given in the specification.  
XX  
SQ Sequence 79 AA;

Query Match 32.4%; Score 79; DB 6; Length 79;  
Best Local Similarity 100.0%; Pred. No. 8.5e-73;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDLGTLGSCLOQFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGR 113  
Db 1 QRAHGQDLGTLGSCLOQFTTTPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPAPITGR 60

QY 114 ALEPYISRCTVCEGPAIAI 132  
Db 61 ALEPYISRCTVCEGPAIAI 79

RESULT 33  
AAU75599  
ID AAU75599 standard; protein; 65 AA.

XX AAU75599;

XX 08-MAY-2002 (first entry)

XX Human type IV collagen alpha 3 chain mutant, Tum-4.

XX Human; type IV collagen alpha 3 chain; cyostatic; antiangiogenic;  
XX non-Goodpasture fragment; alpha3(IV)NCI domain; alphavbeta3 integrin;  
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX Tumstatin; angiogenesis; tumour; mutein; mutant.

XX Homo sapiens.

XX WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

XX 04-APR-2000; 2000US-00543371.

XX 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

XX A non-Goodpasture fragment of alpha3(IV)NCI domain used in detecting and  
XX treating disorders involving angiogenesis.

XX Example 36; Page; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NCI  
CC domain, having one or more of the characteristics selected from: (a) the  
CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit

CC proliferation of endothelial cells; and (c) the ability to cause  
CC apoptosis of endothelial cells. Also described are the following: (1) use  
CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
CC analogue or allelic variant in the preparation of a medicament for  
CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
CC where the angiogenesis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; or (b) by  
CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
CC endothelial cell apoptosis is mediated by one or more endothelial cell  
CC integrins or one or more endothelial cell integrin subunits; (2) use of  
CC an antibody or peptide that specifically binds the alpha1, alpha2,  
CC alpha3, alpha4, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
CC preparation of a medicament for inhibiting angiogenesis or cell  
CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
CC of a medicament for treating a proliferative disease in a vertebrate,  
CC where the disease is characterised by angiogenesis that is mediated by  
CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
CC the presence of a medicament for promoting angiogenesis in a tissue; and  
CC (5) use of integrins in the preparation of a medicament for promoting or  
CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
CC or allelic variants are useful in the preparation of a medicament for  
CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
CC the angiogenesis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits; or by promoting or  
CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
CC cell apoptosis is mediated by one or more endothelial cell integrins or  
CC one or more endothelial cell integrin subunits. The medicament is useful  
CC in inhibiting tumour growth and for the regression of an established  
CC tumour. The present sequence represents the amino acid sequence of human  
CC type IV collagen alpha 3 chain mutant, Tum-4, which consists of residues  
CC 181-244 of Tumstatin. Note: The present sequence is not shown in the  
CC specification but is derived from the wild type human Tumstatin sequence  
CC given in figure 18A (see AAU5589)

SQ Sequence 65 AA;

Query Match 26.6%; Score 65; DB 5; Length 65;  
Best Local Similarity 100.0%; Pred. No. 1.6e-58;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 179 LEEFRASPFLFCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQV 238  
Db 1 LEEFRASPFLFCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQV 60  
Qy 239 CMKKR 243  
Db 61 CMKKR 65

RESULT 34

ADA20263  
XX ADA20263 standard; protein; 64 AA.

AC ADA20263;

XX 20-NOV-2003 (first entry)

DE Human tumstatin deletion protein tum-4 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cyostatic; gene therapy; alpha 3 chain; tumstatin; human; tum-4.

OS Homo sapiens.

XX WO20003059257-A2.

XX

PD 24-JUL-2003.  
XX 20-DEC-2002; 2002WO-US040938.  
PF 21-DEC-2001; 2001US-00032221.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
PA Kalluri R;  
PI WPI; 2003-587256/55.  
XX N-PSDB; ADA20224.  
DR New peptide, useful for preparing a composition for inhibiting tumour  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
PT Claim 94; SEQ ID NO 25; 240pp; English.  
PS This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of tum-4, an abridged form of the "tumstatin" protein of  
CC the invention which was derived from the amino acid sequence of the alpha  
CC 3 chain of human type IV collagen. Note: This sequence (Seq ID25) does  
CC not appear in the specification but was created by the indexer from  
CC information given in the specification.

SQ Sequence 64 AA;

Query Match 26.2%; Score 64; DB 6; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.7e-57;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 EFRASPFLFCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 240  
Db 1 EFRASPFLFCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAGELEKIISRCQVCM 60  
Qy 241 KKRH 244  
Db 61 KKRH 64

RESULT 35

AA95920

ID AA95920 standard; protein; 68 AA.

AC AA95920;

XX 20-NOV-2000 (first entry)

DE Human Goodpasture antigen Deltaii/IV/V.

XX Goodpasture antigen; GPdeltaii/IV/V; human; GPBP;  
KW Goodpasture antigen binding protein; autoimmune disease; apoptosis;  
KW cancer; tumour; therapy.

OS Homo sapiens.

XX WO2000050607-A2.

XX

```

PD 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-IB000324.
XX
XX 24-FEB-1999; 99US-0121483P.
XX
XX (SAUS/) SAUS J.
XX
XX Saus J;
XX
XX WPI; 2000-572094/53.
XX
XX N-PSDB; AAA50369.
XX
XX Novel Goodpasture antigen binding proteins useful for diagnosing and
XX treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX
XX Claim 36; Page 154; 158pp; English.
XX
XX The present sequence is that of human recombinant Goodpasture antigen
XX (GP) DeltaIII/IV/V, i.e. an alternative form of human GP resulting from
XX splicing out of exons III, IV and V. The recombinant protein was
XX expressed in bacterial pellets using modified vector pET15b carrying
XX GPDeltaIII/IV/V cDNA (see AAA50369). The provides Goodpasture antigen
XX binding proteins (GPBs, see AA95900-11), which bind to and
XX phosphorylate the unique N-terminal region of human GP, and which are
XX highly expressed in several autoimmune conditions. Claimed methods for
XX treating an autoimmune disorder, cell apoptosis or a tumour involve
XX modifying the expression or activity of GPBP, especially using a GP-
XX derived peptide, such as GPDeltaIII/IV/V
XX
XX Sequence 68 AA;
XX
Query Match 25.0%; Score 61; DB 3; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-54;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GLGKRGDSGSPATWTRGFVTRHQSQTATPSCPEGTPLVSGFSLFVQGNQRAHGQD 60
DB 1 GLGKRGDSGSPATWTRGFVTRHQSQTATPSCPEGTPLVSGFSLFVQGNQRAHGQD 60
QY 61 L 61
DB 61 L 61
XX
RESULT 36
ABG79210
ID ABG79210 standard; protein; 68 AA.
XX
XX ABG79210;
XX
XX 15-NOV-2002 (first entry)
XX
XX Human GP protein isoform GPDeltaIII/IV/V.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NC1; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX
XX Homo sapiens.
XX
XX WO200261430-A2.
XX
XX 08-AUG-2002.
XX
XX 31-JAN-2002; 2002WO-EP001010.
XX
XX 31-JAN-2001; 2001US-0265249P.
XX
XX (SAUS/) SAUS J.
XX

```

```

XX
XX Saus J;
XX
XX WPI; 2002-619280/66.
XX
XX N-PSDB; ABS64493.
XX
XX Identifying candidate compounds for treating autoimmune conditions, e.g.
XX Goodpasture syndrome or lupus, comprises identifying compounds that
XX reduce phosphorylation of, or formation of conformational isomers of,
XX target proteins.
XX
XX Example 3; Page 202; 217pp; English.
XX
XX The invention relates to identifying candidate compounds to treat an
XX autoimmune condition by identifying compounds that reduce phosphorylation
XX of a first target protein (I) (which is selected from Goodpasture antigen
XX binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NC1)
XX domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-
XX Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising
XX Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of
XX conformational isomers of the second target protein (II) (selected from
XX an alpha3 type IV collagen NC1 domain polypeptide and myelin basic
XX protein, MBP). Also included are (i) an isolated type IV collagen alpha3
XX NC1 domain conformational isomer, which has an amino acid sequence
XX identical to the wild type alpha3 type IV collagen NC1 domain, is
XX stabilised by disulphide bonds, and has a molecular weight in a non-
XX reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in
XX a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated
XX type IV collagen alpha3 NC1 domain. The human gene for GPBP is located on
XX chromosome 5q13. The method is useful for treating autoimmune conditions,
XX such as Goodpasture syndrome, multiple sclerosis, systemic and cutaneous
XX lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present
XX sequence represents an alpha3 type IV collagen non-collagenous (NC1)
XX domain (also known as the GP antigen) or an MBP isoform
XX
XX Sequence 68 AA;
XX
Query Match 25.0%; Score 61; DB 5; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-54;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GLGKRGDSGSPATWTRGFVTRHQSQTATPSCPEGTPLVSGFSLFVQGNQRAHGQD 60
DB 1 GLGKRGDSGSPATWTRGFVTRHQSQTATPSCPEGTPLVSGFSLFVQGNQRAHGQD 60
QY 61 L 61
DB 61 L 61
XX
RESULT 37
AA95919
ID AA95919 standard; protein; 72 AA.
XX
XX AA95919;
XX
XX 20-NOV-2000 (first entry)
XX
XX Human Goodpasture antigen DeltaIII.
XX
XX Goodpasture antigen; GPDeltaIII; human; GPBP;
XX Goodpasture antigen binding protein; autoimmune disease; apoptosis;
XX cancer; tumour; therapy.
XX
XX Homo sapiens.
XX
XX WO200050607-A2.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-IB000324.
XX
XX 24-FEB-1999; 99US-0121483P.
XX

```

```

XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX DR WPI: 2000-572094/53.
XX DR N-PSDB; AAA50368.
XX PT Novel Goodpasture antigen binding proteins useful for diagnosing and
XX PT treating autoimmune disorders, tumor, and preventing cell apoptosis.
XX PS Claim 36; Page 153; 158pp; English.
XX CC The present sequence is that of human recombinant Goodpasture antigen
XX CC (GP) DeltailI/V, i.e. an alternative form of human GP resulting from
XX CC splicing out of exon III. The recombinant protein, lacking the Met-1
XX CC residue, was expressed in bacterial pellets using modified vector pR15b
XX CC carrying GPdeltailI cDNA (see AAA50368). The provides novel Goodpasture
XX CC antigen binding proteins (GPBPs, see AAY95900-11), which bind to and
XX CC phosphorylate the unique N-terminal region of human GP, and which are
XX CC highly expressed in several autoimmune conditions. Claimed methods for
XX CC treating an autoimmune disorder, cell apoptosis or a tumour involve
XX CC modifying the expression or activity of GPBP, especially using a GP-
XX CC derived peptide, such as GPdeltailI
XX SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 3; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.3e-54;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTPSCPGTVPVLYSGFSFLVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTTPSCPGTVPVLYSGFSFLVQGNQRAHQD 60

QY 61 L 61
DB 61 L 61

RESULT 38
AAY95921
ID AAY95921 standard; protein; 72 AA.
XX AC AAY95921;
XX DT 20-NOV-2000 (first entry)
XX DE Human Goodpasture antigen DeltailI/V.
XX KW Goodpasture antigen, GPdeltailI/V; human; GPBP;
XX KW goodpasture antigen binding protein; autoimmune disease; apoptosis;
XX KW cancer; tumour; therapy.
XX OS Homo sapiens.
XX PN WO2000050607-A2.
XX PD 31-AUG-2000.
XX PF 24-FEB-2000; 2000WO-IB000324.
XX PR 24-FEB-1999; 99US-0121483P.
XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX DR WPI: 2000-572094/53.
XX DR N-PSDB; AAA50370.
XX PT Novel Goodpasture antigen binding proteins useful for diagnosing and
XX PT treating autoimmune disorders, tumor, and preventing cell apoptosis.

```

```

XX PS Claim 36; Page 155; 158pp; English.
XX CC The present sequence is that of human recombinant Goodpasture antigen
XX CC (GP) DeltailI/V, i.e. an alternative form of human GP resulting from
XX CC splicing out of exons III and V. The recombinant protein was expressed in
XX CC bacterial pellets using modified vector pR15b carrying GPdeltailI/IV/V
XX CC cDNA (see AAA50369). The provides Goodpasture antigen binding proteins
XX CC (GPBPs, see AAY95900-11), which bind to and phosphorylate the unique N-
XX CC terminal region of human GP, and which are highly expressed in several
XX CC autoimmune conditions. Claimed methods for treating an autoimmune
XX CC disorder, cell apoptosis or a tumour involve modifying the expression or
XX CC activity of GPBP, especially using a GP-derived peptide, such as
XX CC GPdeltailI/V
XX SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 3; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.3e-54;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFVTRHSQTTPSCPGTVPVLYSGFSFLVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGFVTRHSQTTPSCPGTVPVLYSGFSFLVQGNQRAHQD 60

QY 61 L 61
DB 61 L 61

RESULT 39
ABG79209
ID ABG79209 standard; protein; 72 AA.
XX AC ABG79209;
XX DT 15-NOV-2002 (first entry)
XX DE Human GP protein isoform GPdeltailI.
XX KW Goodpasture antigen binding protein; Goodpasture syndrome;
XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX KW autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX KW pemphigoid; lichen planus; human.
XX OS Homo sapiens.
XX PN WO200261430-A2.
XX PD 08-AUG-2002.
XX PF 31-JAN-2002; 2002WO-EP001010.
XX PR 31-JAN-2001; 2001US-0265249P.
XX PA (SAUS/) SAUS J.
XX PI Saus J;
XX DR WPI: 2002-619280/66.
XX DR N-PSDB; ABS64492.
XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.
XX PT Goodpasture syndrome or lupus, comprises identifying compounds that
XX PT reduce phosphorylation of, or formation of conformational isomers of,
XX PT target proteins.
XX PS Example 3; Page 201; 217pp; English.
XX CC The invention relates to identifying candidate compounds to treat an
XX CC autoimmune condition by identifying compounds that reduce phosphorylation

```

CC of a first target protein (I) (which is selected from Goodpasture antigen  
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)  
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-  
 CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
 CC Lys-Arg-Pro-Gln-Arg-His-Gly), or reduce formation of  
 CC conformational isomers of the second target protein (II) (selected from  
 CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
 CC NCI domain conformational isomer, which has an amino acid sequence  
 CC identical to the wild type alpha3 type IV collagen NCI domain, is  
 CC stabilised by disulphide bonds, and has a molecular weight in a non-  
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
 CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)  
 CC domain (also known as the GP antigen) or an MBP isoform  
 XX  
 SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKRGDGS PATTTTRGTFVTRHSQT TAIPSCPEGT VPLYSGFSFLVQGNQRAHQD 60  
 DB 1 GLKRGDGS PATTTTRGTFVTRHSQT TAIPSCPEGT VPLYSGFSFLVQGNQRAHQD 60  
 QY 61 L 61  
 DB 61 L 61

## RESULT 40

ABG79211  
 ID ABG79211 standard; protein; 72 AA.

XX AC ABG79211;

XX DT 15-NOV-2002 (first entry)

XX DE Human GP protein isoform GPdeltaII/IV.

XX KW Goodpasture antigen binding protein; Goodpasture syndrome;  
 XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
 XX autoimmune condition; phosphorylation; myelin basic protein; MBP;  
 XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;  
 XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
 XX pemphigoid; lichen planus; human.

XX OS Homo sapiens.

XX PN WO200261430-A2.

XX PD 08-AUG-2002.

XX PF 31-JAN-2002; 2002WO-EP001010.

XX PR 31-JAN-2001; 2001US-0265249P.

XX PA (SAUS/) SAUS J.

XX PI Saus J;

XX DR WPI; 2002-619280/56.

XX DR N-PSDB; ABS64494.

XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
 PT Goodpasture syndrome or lupus, comprises identifying compounds that  
 PT reduce phosphorylation of, or formation of conformational isomers of,  
 PT target proteins.

XX PS Disclosure; Page 203-204; 217pp; English.  
 XX CC The invention relates to identifying candidate compounds to treat an  
 CC autoimmune condition by identifying compounds that reduce phosphorylation  
 CC of a first target protein (I) (which is selected from Goodpasture antigen  
 CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)  
 CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-  
 CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
 CC Lys-Arg-Pro-Gln-Arg-His-Gly), or reduce formation of  
 CC conformational isomers of the second target protein (II) (selected from  
 CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
 CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
 CC NCI domain conformational isomer, which has an amino acid sequence  
 CC identical to the wild type alpha3 type IV collagen NCI domain, is  
 CC stabilised by disulphide bonds, and has a molecular weight in a non-  
 CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
 CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
 CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
 CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
 CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
 CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)  
 CC domain (also known as the GP antigen) or an MBP isoform  
 XX  
 SQ Sequence 72 AA;

Query Match 25.0%; Score 61; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-54;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKRGDGS PATTTTRGTFVTRHSQT TAIPSCPEGT VPLYSGFSFLVQGNQRAHQD 60  
 DB 1 GLKRGDGS PATTTTRGTFVTRHSQT TAIPSCPEGT VPLYSGFSFLVQGNQRAHQD 60  
 QY 61 L 61  
 DB 61 L 61

## RESULT 41

ADD47061  
 ID ADD47061 standard; protein; 230 AA.

XX AC ADD47061;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein AAB72238, SEQ ID NO 12749.

XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; AAB72238.

PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC the sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 230 AA;

Query Match 16.0%; Score 39; DB 7; Length 230;

Best Local Similarity 100.0%; Pred. NO. 2.5e-31;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYSNSYFWLASLNPERMFRKPIPTVKAG 227

Db 175 ECHGRGTCNYSNSYFWLASLNPERMFRKPIPTVKAG 213

RESULT 42

AAV44171

ID AAY44171 standard; protein; 471 AA.

XX AAY44171;

XX 01-FEB-2000 (first entry)

XX Bovine type IV collagen alpha3 chain protein.

XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;  
XX Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.

XX Bos taurus.

XX US5973120-A.

XX 26-OCT-1999.

XX 07-MAR-1995; 95US-00399889.

XX 30-NOV-1990; 90US-00621091.

XX (UYA ) UNIV YALE.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Reenders ST, Morrison KE;

XX WPI; 1999-610317/52.

DR N-PSDB; AA228774.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for  
PT diagnosis and treatment of Goodpasture syndrome.

PS Claim 1; Col 31-34; 27pp; English.

XX  
CC This sequence represents a recombinant bovine alpha3 chain of type IV  
CC collagen polypeptide. The sequence corresponds to the 238 amino acids of  
CC the C-terminal end of the triple helical domain and all 233 amino acids  
CC of the C-terminal non-collagenous domain. Alpha3 chain collagen  
CC polypeptides are useful for detecting Goodpasture antibodies in blood or  
CC tissue from a human patient and for treating Goodpasture syndrome,  
CC especially by neutralising the antibodies in the blood. The polypeptides  
CC also have a nephrotrophic activity

XX Sequence 471 AA;

Query Match 16.0%; Score 39; DB 2; Length 471;

Best Local Similarity 100.0%; Pred. NO. 4.7e-31;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCP 145

Db 334 MAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCP 372

RESULT 43

AAV56783

ID AAY56783 standard; protein; 471 AA.

XX AAY56783;

XX 27-MAR-2000 (first entry)

XX Bovine alpha3 type IV collagen.

XX Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.

XX Bos sp.

XX US6007980-A.

XX 28-DEC-1999.

XX 07-OCT-1998; 98US-00167364.

XX 30-NOV-1990; 90US-00621091.

XX 07-MAR-1995; 95US-00399889.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX (UYA ) UNIV YALE.

XX Hudson BG, Reenders ST, Morrison KE;

XX WPI; 2000-096371/08.

XX N-PSDB; AA246728.

XX Diagnosing and treating Goodpasture syndrome using a peptide derived from  
PT type IV collagen.

XX Disclosure; Col 19-24; 26pp; English.

XX The invention provides a method of detecting Goodpasture antibodies in  
CC the fluid of a patient by contacting it with a peptide comprising at most  
CC 218 amino acids of the human alpha3 chain type IV collagen that contains  
CC the fragment shown in AAY56785. The methods are useful for the diagnosis  
CC and treatment of Goodpasture syndrome. The present sequence represents  
CC the bovine alpha3 chain of type IV collagen

XX Sequence 471 AA;

Query Match 16.0%; Score 39; DB 3; Length 471;



Best Local Similarity 100.0%; Pred. No. 4.7e-31; Mismatches 0; Indels 0; Gaps 0;  
Matches 39; Conservative 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 145  
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 372

RESULT 44  
AAE09483  
ID AAE09483 standard; protein; 471 AA.  
XX AAE09483;  
XX  
XX  
DT 19-NOV-2001 (first entry)  
XX  
DE Bovine alpha-3 chain of type IV collagen protein.  
XX  
XX Bovine; alpha-3 chain; type IV collagen; immunosuppressive; therapy;  
KW Goodpasture syndrome.  
XX  
XX Bos taurus.  
XX  
XX US6277558-B1.  
XX  
XX 21-AUG-2001.  
XX  
XX 12-NOV-1999; 99US-00439897.  
XX  
XX 30-NOV-1990; 90US-00621091.  
PR 07-MAR-1995; 95US-00399889.  
PR 07-OCT-1998; 98US-00167364.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
XX Hudson BG;  
XX  
XX WPI; 2001-540401/60.  
DR N-PSDB; AADI6399.  
XX  
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting  
PT Goodpasture antibodies from bodily fluid/tissue from patient or for  
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with  
PT the polypeptide.  
XX  
XX Example 4; Col 33-36; 46pp; English.  
XX  
XX The invention relates to a method for detecting Goodpasture antibodies  
CC from a bodily fluid or tissue of a patient. The method comprises  
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)  
CC collagen polypeptide that contains a conformational epitope for the  
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for  
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a  
CC patient, and for treating Goodpasture syndrome in a patient. The present  
CC sequence is bovine alpha-3 chain of type IV collagen protein  
XX  
SQ Sequence 471 AA;  
Query Match 16.0%; Score 39; DB 4; Length 471;  
Best Local Similarity 100.0%; Pred. No. 4.7e-31;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 145  
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCP 372

RESULT 45  
ABG79213  
ID ABG79213 standard; peptide; 72 AA.  
XX  
XX  
AC ABG79213;  
XX

15-NOV-2002 (first entry)  
XX Human GP protein isoform GPdeltaIII phosphorylation site region.  
DE  
XX Goodpasture antigen binding protein; Goodpasture syndrome;  
KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
KW autoimmune condition; phosphorylation; myelin basic protein; MBP;  
KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;  
KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
KW pemphigoid; lichen planus; human.  
XX  
XX Homo sapiens.  
XX  
XX WO200261430-A2.  
XX  
XX 08-AUG-2002.  
XX  
XX 31-JAN-2002; 2002WO-EP001010.  
XX  
XX 31-JAN-2001; 2001US-0265249P.  
XX  
XX (SAUS/) SAUS J.  
XX  
XX Saus J;  
XX  
XX WPI; 2002-619280/66.  
XX  
XX Identifying candidate compounds for treating autoimmune conditions, e.g.  
PT Goodpasture syndrome or lupus, comprises identifying compounds that  
PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.  
XX  
XX Example 3; Fig 17; 217pp; English.  
XX  
XX The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (I) (which is selected from Goodpasture antigen  
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Trp-Thr-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of  
CC conformational isomers of the second target protein (II) (selected from  
CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC NCI domain conformational isomer, which has an amino acid sequence  
CC identical to the wild type alpha3 type IV collagen NCI domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated  
CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)  
CC domain (also known as the GP antigen) or an MBP phosphorylation site  
XX  
SQ Sequence 72 AA;  
Query Match 15.2%; Score 37; DB 5; Length 72;  
Best Local Similarity 100.0%; Pred. No. 9.7e-30;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGFVTRHSQTTAIPSCPEG 37  
DB 1 GLKGRGDSGPATWTRGFVTRHSQTTAIPSCPEG 37

RESULT 46  
AAE09503  
ID AAE09503 standard; peptide; 36 AA.  
XX  
XX AAE09503;  
XX

```
DT 19-NOV-2001 (first entry)
XX Human C8 alpha-3 peptide to construct alpha1/alpha3(IV)NC1 protein.
DE
XX
XX
XX Human alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome; C8 alpha-3 peptide.
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
XX 07-MAR-1995; 95US-00399889.
XX 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WPI; 2001-540401/60.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
XX Goodpasture antibodies from bodily fluid/tissue from patient or for
XX treating Goodpasture syndrome by contacting bodily fluid or tissue with
XX the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
XX
XX The invention relates to a method for detecting Goodpasture antibodies
XX from a bodily fluid or tissue of a patient. The method comprises
XX contacting the bodily fluid or tissue with alpha-3 chain type (IV)
XX collagen polypeptide that contains a conformational epitope for the
XX Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
XX detecting Goodpasture antibodies from a bodily fluid or tissue from a
XX patient, and for treating Goodpasture syndrome in a patient. The present
XX sequence is human alpha chain peptide used for constructing human
XX alpha1/alpha3 (IV) NC1 fusion protein
XX
XX Sequence 36 AA;
XX
XX Query Match 14.8%; Score 36; DB 4; Length 36;
XX Best Local Similarity 100.0%; Pred. No. 5.5e-29;
XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 209 SLNPFMRFKPISTVKAGELEKIIIRQCQVCMKRRH 244
XX |||||||
XX 1 SLNPFMRFKPISTVKAGELEKIIIRQCQVCMKRRH 36
XX
XX RESULT 47
XX AAR79163
XX ID AAR79163 standard; protein; 471 AA.
XX
XX AC AAR79163;
XX
XX 22-DEC-1995 (first entry)
XX
XX Partial sequence of bovine alpha 3 chain of type IV collagen.
XX
XX Type IV collagen; alpha 3 chain.
XX
XX Bos taurus.
XX
XX US5424408-A.
XX
XX 13-JUN-1995.
XX
XX 30-NOV-1990; 90US-00621091.
XX
```

```
PR 30-NOV-1990; 90US-00621091.
XX
XX (UYJA ) UNIV YALE.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Morrison KE, Reenders ST, Hudson BG;
XX
XX WPI; 1995-262631/34.
XX N-PSDB; AAQ96290.
XX
XX CDNA's encoding human or bovine alpha-3 type 4 collagen peptide(s) -
XX useful for detection and therapeutic removal of antibodies associated
XX with Goodpasture syndrome.
XX
XX Disclosure; Col 5-8; 33pp; English.
XX
XX Using the PCR with primers derived from each end of the known 27 AA
XX residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp bovine
XX genomic fragment was amplified. This fragment was then used to a bovine
XX lens cDNA library and a 1.5 kb partial cDNA clone was obt'd. (clone
XX KM15). This encodes 238 residues of the triple helical collagenous
XX domain and all 233 residues of the C-terminal non-collagenous (NC1)
XX domain of the alpha 3 (IV) chain. An isolated and substantially pure nt.
XX having the sequence in AAQ96290 is claimed
XX
XX Sequence 471 AA;
XX
XX Query Match 12.3%; Score 30; DB 2; Length 471;
XX Best Local Similarity 100.0%; Pred. No. 8.1e-22;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 116 EPYISRCTVCEGPAIAIAVHSQTTDIPPCP 145
XX |||||||
XX 343 EPYISRCTVCEGPAIAIAVHSQTTDIPPCP 372
XX
XX RESULT 48
XX AAE09501
XX ID AAE09501 standard; peptide; 26 AA.
XX
XX AC AAE09501;
XX
XX 19-NOV-2001 (first entry)
XX
XX Human C7 alpha-3 peptide to construct alpha1/alpha3(IV)NC1 protein.
XX
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
XX Goodpasture syndrome; C7 alpha-3 peptide.
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
XX 07-MAR-1995; 95US-00399889.
XX 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WPI; 2001-540401/60.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
XX Goodpasture antibodies from bodily fluid/tissue from patient or for
XX treating Goodpasture syndrome by contacting bodily fluid or tissue with
XX the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
XX
```

XX The invention relates to a method for detecting Goodpasture antibodies  
CC from a bodily fluid or tissue of a patient. The method comprises  
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)  
CC collagen polypeptide that contains a conformational epitope for the  
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for  
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a  
CC patient, and for treating Goodpasture syndrome in a patient. The present  
CC sequence is human alpha chain peptide used for constructing human  
CC alpha1/alpha3 (IV) NC1 fusion protein  
XX  
SQ Sequence 26 AA;  
Query Match 10.7%; Score 26; DB 4; Length 26;  
Best Local Similarity 100.0%; Pred. No. 7.5e-19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGFVTRHS 26  
DB 1 GLKGRGDSGSPATWTRGFVTRHS 26  
RESULT 49  
ADA20238  
ID ADA20238 standard; peptide; 27 AA.  
XX  
AC ADA20238;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE T8 peptide related to human type IV collagen alpha and angiogenesis.  
XX  
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytostatic; gene therapy; T8 peptide; tumstatin; human;  
XX type IV collagen alpha 3 chain; mutant; mutein.  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX  
PH Key Location/Qualifiers  
FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"  
FT  
XX  
XX WO2003059257-A2.  
XX  
XX 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
XX  
XX 21-DEC-2001; 2001US-00032221.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX  
XX WPI; 2003-587256/55.  
XX  
XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX Claim 62; Page 45; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC

CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is the amino acid sequence of the T8 peptide of the invention,  
CC derived from the amino acid sequence of tumstatin, which in turn was  
CC derived from the amino acid sequence of human type IV collagen alpha 3  
CC chain.  
XX  
SQ Sequence 27 AA;  
Query Match 10.7%; Score 26; DB 6; Length 27;  
Best Local Similarity 100.0%; Pred. No. 7.7e-19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 QRFTTWPFLFCNVNDVCFASRNDYS 94  
DB 2 QRFTTWPFLFCNVNDVCFASRNDYS 27  
RESULT 50  
ADA20236  
ID ADA20236 standard; peptide; 25 AA.  
XX  
AC ADA20236;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE T7 peptide related to human type IV collagen alpha and angiogenesis.  
XX  
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytostatic; gene therapy; T7 peptide; tumstatin; human;  
XX type IV collagen alpha 3 chain.  
XX Homo sapiens.  
XX  
XX WO2003059257-A2.  
XX  
XX 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
XX  
XX 21-DEC-2001; 2001US-00032221.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX  
XX WPI; 2003-587256/55.  
XX N-PSDB; ADA20224.  
XX  
XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX Claim 53; Page 45; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is the amino acid sequence of the 17 peptide of the invention,  
CC derived from the amino acid sequence of human type IV collagen alpha 3  
CC chain.

XX SQ Sequence 25 AA;

Query Match 10.2%; Score 25; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 7.7e-18;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMTPLFCNVNDVCNFSARNDSYWL 97  
Db 1 TMTPLFCNVNDVCNFSARNDSYWL 25  
|||||

RESULT 51  
ADCI17661  
ID ADCI17661 standard; peptide; 22 AA.

XX AC ADCI17661;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:266.

XX DE crystallised NC1 domain hexamer of type IV collagen;  
XX KW angiogenesis inhibitor; angiogenesis-mediated disease;  
XX KW tumour metastasis inhibitor; tumour growth inhibitor;  
XX KW endothelial cell interaction inhibitor;  
XX KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
XX KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
XX KW endothelial cell adhesion inhibitor;  
XX KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
XX KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
XX KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV.) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
XX PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
XX PT basal lamina membrane formation in cell or tissue development.

XX PS Claim 57; SEQ ID NO 266; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
XX CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a

CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
CC invention.

XX SQ Sequence 22 AA;

Query Match 9.0%; Score 22; DB 7; Length 22;  
Best Local Similarity 100.0%; Pred. No. 8.2e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTMPFLFCNVNDVCNFSARNSD 92

Db 1 FTTMPFLFCNVNDVCNFSARNSD 22  
|||||

RESULT 52

ADCI17414

ID ADCI17414 standard; peptide; 22 AA.

XX AC ADCI17414;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:15.

XX DE crystallised NC1 domain hexamer of type IV collagen;  
XX KW angiogenesis inhibitor; angiogenesis-mediated disease;  
XX KW tumour metastasis inhibitor; tumour growth inhibitor;  
XX KW endothelial cell interaction inhibitor;  
XX KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
XX KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
XX KW endothelial cell adhesion inhibitor;  
XX KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
XX KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
XX KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

FT Misc-difference 1..5

FT Misc-difference 18..22

XX WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

```
PR 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAWOORTHY M.
PA (HUDS/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glioma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.
XX
XX Claim 11; SEQ ID NO 15; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (5) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of an NCI
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and
CC antitumor activities, and can be used as an inhibitor of angiogenesis,
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glioma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX SQ Sequence 22 AA;
XX
XX Query Match 9.0%; Score 22; DB 7; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 8.2e-15;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 181 EFRASPFLCHGRTGNYNS 202
XX Db 1 EFRASPFLCHGRTGNYNS 22
XX
XX RESULT 53
XX AAY95912
XX ID AAY95912 standard; peptide; 21 AA.
XX AC AAY95912;
XX XX
XX 20-NOV-2000 (first entry)
XX DT
XX DE Human Goodpasture antigen N-terminal peptide GPpepl.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome; antigen;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX OS
XX Homo sapiens.
XX XX
XX WO2000261430-A2.
XX FN
XX 08-AUG-2002.
XX PD
XX 31-JAN-2002; 2002MO-EP001010.
XX PF
XX 31-JAN-2001; 2001US-0265249P.
XX PR
XX (SAUS/) SAUS J.
XX PA
XX Saus J;
XX PI
XX WPI; 2002-619280/66.
XX DR
XX XX
```

```
PF 24-FEB-2000; 2000WO-IB000324.
XX
XX 24-FEB-1999; 99US-0121483P.
XX
XX (SAUS/) SAUS J.
XX PA
XX Saus J;
XX PI
XX WPI; 2000-572094/53.
XX DR
XX
XX Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis.
PT
XX Example 1; Page 21; 158pp; English.
XX
XX The present sequence is that of GPpepl, the N-terminal 21 amino acids of
CC human Goodpasture antigen (GP). The peptide was used to search for
CC proteins that interacted with it, leading to the identification of human
CC Goodpasture binding protein (GPBP, see AAY95900), a novel
CC serine/threonine kinase that specifically binds to and phosphorylates
CC GPpepl. GPpepl was also used to characterise the phosphorylation activity
CC of GPBP. The invention provides nucleic acids (see AA50341-53) encoding
CC GPBP, recombinant vectors, host cells, encoded polypeptides (see AAY95900
CC -1) and antibodies. It also provides methods for detecting the presence
CC of an autoimmune condition or apoptosis by detecting an increase in GPBP
CC expression, and methods for treating an autoimmune disorder, apoptosis or
CC a tumour by modifying GPBP expression or activity, especially using a GP-
XX derived peptide
XX
XX SQ Sequence 21 AA;
XX
XX Query Match 8.6%; Score 21; DB 3; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-14;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 KGRGDSGSPATWTRGFVFT 23
XX Db 1 KGRGDSGSPATWTRGFVFT 21
XX
XX RESULT 54
XX ABG79202
XX ID ABG79202 standard; peptide; 21 AA.
XX XX
XX AC ABG79202;
XX XX
XX 15-NOV-2002 (first entry)
XX DT
XX DE Human Goodpasture protein peptide, GPpepl.
XX
XX Goodpasture antigen binding protein; Goodpasture syndrome; antigen;
XX chromosome 5q13; neuroprotective; dermatological; immunosuppressive;
XX autoimmune condition; phosphorylation; myelin basic protein; MBP;
XX alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;
XX systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;
XX pemphigoid; lichen planus; human.
XX OS
XX Homo sapiens.
XX XX
XX WO200261430-A2.
XX FN
XX 08-AUG-2002.
XX PD
XX 31-JAN-2002; 2002MO-EP001010.
XX PF
XX 31-JAN-2001; 2001US-0265249P.
XX PR
XX (SAUS/) SAUS J.
XX PA
XX Saus J;
XX PI
XX WPI; 2002-619280/66.
XX DR
XX XX
```

PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
PT Goodpasture syndrome or lupus, comprises identifying compounds that  
PT reduce phosphorylation of, or formation of conformational isomers of,  
PT target proteins.

PS Claim 1; Page 26; 217pp; English.

XX The invention relates to identifying candidate compounds to treat an  
CC autoimmune condition by identifying compounds that reduce phosphorylation  
CC of a first target protein (1) (which is selected from Goodpasture antigen  
CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCL)  
CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly- Asp-Ser-Gly-Ser-Pro-  
CC Ala-Thr-Trip-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of  
CC conformational isomers of the second target protein (11) (selected from  
CC an alpha3 type IV collagen NCL domain polypeptide and myelin basic  
CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
CC NCL domain conformational isomer, which has an amino acid sequence  
CC identical to the wild type alpha3 type IV collagen NCL domain, is  
CC stabilised by disulphide bonds, and has a molecular weight in a non-  
CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated  
CC type IV collagen alpha3 NCL domain. The human gene for GPBP is located on  
CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
CC sequence represents an alpha3 type IV collagen non-collagenous (NCL)  
CC domain (also known as the GP antigen) peptide antigen

XX Sequence 21 AA;

Query Match 8.6%; Score 21; DB 5; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.3e-14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGRGDSGSPATWTRGFVPT 23  
DB 1 KGRGDSGSPATWTRGFVPT 21  
|||||

RESULT 55

ADCL7642  
ID ADC17642 standard; peptide; 21 AA.

XX AC ADC17642;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NCL domain related peptide SEQ ID NO:247.

XX crystallised NCL domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiapoptotic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO2003012122-A2.

XX XX 13-FEB-2003.

XX XX 26-JUL-2002; 2002WO-US023763.

XX XX 27-JUL-2001; 2001US-0308523P.

XX XX 29-OCT-2001; 2001US-0351289P.

XX XX 22-MAR-2002; 2002US-0366854P.

PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

PI Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 247; 168pp; English.

XX The present invention describes a crystallised NCL domain hexamer of type  
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NCL  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NCL domain hexamer of type IV collagen (I) has cytostatic,  
CC antiapoptotic, antianaemic, ophthalmological, antiarteriosclerotic, and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
XX invention.

XX Sequence 21 AA;

Query Match 9.6%; Score 21; DB 7; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.3e-14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 FWLASLNPFRMFKPIPISTVK 225

DB 1 FWLASLNPFRMFKPIPISTVK 21  
|||||

RESULT 56

AAU75604

ID AAU75604 standard; peptide; 20 AA.

XX AC AAU75604;

XX DT 08-MAY-2002 (first entry)

XX DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T4.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NCL domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutein; mutant.

XX OS Homo sapiens.

XX PN WO200151523-A2.

XX XX 19-JUL-2001.

```

XX PF 08-JAN-2001; 2001WO-US000565.
XX PR
XX PR 07-JAN-2000; 2000US-00479118.
XX PR 04-APR-2000; 2000US-00543371.
XX PR 21-JUL-2000; 2000US-00625191.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Kalluri R;
XX DR WPI; 2002-188037/24.
XX PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and
XX PT treating disorders involving angiogenesis.
XX PS Example 40; Page 133; 205pp; English.
XX CC The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1
XX CC domain, having one or more of the characteristics selected from: (a) the
XX CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit
XX CC proliferation of endothelial cells; and (c) the ability to cause
XX CC apoptosis of endothelial cells. Also described are the following: (1) use
XX CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,
XX CC analogue or allelic variant in the preparation of a medicament for
XX CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,
XX CC where the angiogenesis is mediated by one or more endothelial cell
XX CC integrins or one or more endothelial cell integrin subunits; (2) use of
XX CC an antibody or peptide that specifically binds the alpha1, alpha2,
XX CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the
XX CC preparation of a medicament for inhibiting angiogenesis or cell
XX CC proliferation; (3) use of an inhibitor, such as an antibody, antibody
XX CC fragment or peptide of receptor-mediated angiogenesis in the preparation
XX CC of a medicament for treating a proliferative disease in a vertebrate,
XX CC where the disease is characterised by angiogenesis that is mediated by
XX CC receptors to Arresten, Canstatin or Tumstatin and where the receptors
XX CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one
XX CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in
XX CC the presence of a medicament for promoting angiogenesis in a tissue; and
XX CC (5) use of integrins in the preparation of a medicament for promoting or
XX CC inducing angiogenesis or cell proliferation in a tissue. The fragments
XX CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues
XX CC or allelic variants are useful in the preparation of a medicament for
XX CC treating a disorder involving inhibiting angiogenesis in a tissue, where
XX CC the angiogenesis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits; or by promoting or
XX CC inducing endothelial cell apoptosis in a tissue, where the endothelial
XX CC cell apoptosis is mediated by one or more endothelial cell integrins or
XX CC one or more endothelial cell integrin subunits. The medicament is useful
XX CC in inhibiting tumour growth and for the regression of an established
XX CC tumour. The present sequence represents the amino acid sequence of human
XX CC type IV collagen alpha 3 chain mutant, T4, which consists of residues 84-
XX CC 103 of Tumstatin
XX SQ Sequence 20 AA;

Query Match      8.2%; Score 20; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-13; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 83 DVCNFSARNDSYWLSTPAL 102
   |||||
Db 1 DVCNFSARNDSYWLSTPAL 20

RESULT 57
AAU75602
ID AAU75602 standard; peptide; 20 AA.
XX
XX AC AAU75602;

```

08-MAY-2002 (first entry)

Human type IV collagen alpha 3 chain, Tumstatin, mutant T2.

Human; type IV collagen alpha 3 chain; cytosolic; antiangiogenic;  
non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
Tumstatin; angiogenesis; tumour; mutant; mutant.

Homo sapiens.

WO200151523-A2.

19-JUL-2001.

08-JAN-2001; 2001WO-US000565.

07-JAN-2000; 2000US-00479118.

04-APR-2000; 2000US-00543371.

21-JUL-2000; 2000US-00625191.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

Kalluri R;

WPI; 2002-188037/24.

A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and treating disorders involving angiogenesis.

Example 40; Page 133; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, T4, which consists of residues 84-103 of Tumstatin

Sequence 20 AA;



Query Match 8.2%; Score 20; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 NORAHGQDLTGSLCLORET 72  
DB 1 NORAHGQDLTGSLCLORET 20  
|||||

RESULT 58  
AAU75603  
ID AAU75603 standard; peptide; 20 AA.  
XX AAU75603;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T3.  
XX  
KW Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour; mutant;  
XX  
OS Homo sapiens.  
XX  
PN WO200151523-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 08-JAN-2001; 2001WO-US000565.  
XX  
PR 07-JAN-2000; 2000US-00479118.  
PR 04-APR-2000; 2000US-00543371.  
PR 21-JUL-2000; 2000US-00625191.  
XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
PI Kalluri R;  
XX  
PF WPI; 2002-188037/24.  
XX  
PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
PT treating disorders involving angiogenesis.  
XX  
PS Claim 33; Page 153; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate where the disease is characterised by angiogenesis that is mediated by receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or

inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 3 chain mutant, T3, which consists of residues 69-88 of Tumstatin

XX Sequence 20 AA;  
SQ

Query Match 8.2%; Score 20; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LQRETTMPFLFCNVNVCNF 87  
DB 1 LQRETTMPFLFCNVNVCNF 20  
|||||

RESULT 59  
ADA20267  
ID ADA20267 standard; peptide; 20 AA.  
XX  
AC ADA20267;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human tumstatin deletion peptide T3 amino acid sequence.  
XX  
KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth; metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis; cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T3.  
XX  
OS Homo sapiens.  
XX  
PN WO2003059257-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 20-DEC-2002; 2002WO-US040938.  
XX  
PR 21-DEC-2001; 2001US-00032221.  
XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
PI Kalluri R;  
XX  
DR WPI; 2003-587256/55.  
DR N-PSDB; ADA20224.  
XX  
PT New peptide, useful for preparing a composition for inhibiting tumour growth, angiogenic activity or protein synthesis in a mammalian tissue.  
PS Claim 94; Page 131; 240pp; English.

This invention relates to novel isolated proteins and their fragments with anti-angiogenic properties. The invention also relates to the DNA sequences which encode the novel proteins. A wide variety of diseases are the result of undesirable angiogenesis. The formation of new capillaries from pre-existing vessels is essential for tumour growth and metastasis. Basement membrane organisation is dependent on the assembly of a type IV collagen network which may occur through the C-terminal globular non-collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2 forms are ubiquitously exhibited in human basement membranes. In the present invention, cell surface receptors (in particular integrins) which

CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of peptide T3, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen.  
XX  
XX Sequence 20 AA;

Query Match 8.2%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 LQFTTTPFLFCNVNVCNF 87  
Db 1 LQFTTTPFLFCNVNVCNF 20

RESULT 60  
ADA20266  
ID ADA20266 standard; peptide; 20 AA.

XX ADA20266;  
XX  
XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion peptide T2 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T2.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Example 40; Page 131; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which  
XX specifically bind anti-angiogenic proteins and peptides (in particular  
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX growth, angiogenic activity in mammalian tissue or protein synthesis in

CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of peptide T2, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen.  
XX  
XX Sequence 20 AA;

Query Match 8.2%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 NQRAHQDGLTGLGSCLORET 72  
Db 1 NQRAHQDGLTGLGSCLORET 20

RESULT 61  
ADA20268  
ID ADA20268 standard; peptide; 20 AA.

XX ADA20268;

XX 20-NOV-2003 (first entry)

XX Human tumstatin deletion peptide T4 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T4.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor  
XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 94; Page 131; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
XX with anti-angiogenic properties. The invention also relates to the DNA  
XX sequences which encode the novel proteins. A wide variety of diseases are  
XX the result of undesirable angiogenesis. The formation of new capillaries  
XX from pre-existing vessels is essential for tumour growth and metastasis.  
XX Basement membrane organisation is dependent on the assembly of a type IV  
XX collagen network which may occur through the C-terminal globular non-  
XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX forms are ubiquitously exhibited in human basement membranes. In the  
XX present invention, cell surface receptors (in particular integrins) which  
XX specifically bind anti-angiogenic proteins and peptides (in particular  
XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX sequences of the invention may be useful in gene therapy. The present  
XX sequence is that of peptide T4, an abridged form of the "tumstatin"  
XX protein of the invention which was derived from the amino acid sequence

CC of the alpha 3 chain of human type IV collagen.  
XX Sequence 20 AA;  
SQ

Query Match 8.2%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNFSRNDYSYWLSTPAL 102  
Db 1 DVCNFSRNDYSYWLSTPAL 20  
|||||

RESULT 62  
ADCL17684  
ID ADCL17684 standard; peptide; 20 AA.  
XX  
AC ADCL17684;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
XX Type IV collagen NC1 domain related peptide SEQ ID NO:289.  
XX  
XX crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiapoptotic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
XX Synthetic.  
OS  
OS Homo sapiens.  
XX  
XX WO2003012122-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 26-JUL-2002; 2002WO-US023763.  
XX  
XX 27-JUL-2001; 2001US-0308523P.  
XX  
XX 29-OCT-2001; 2001US-0351289P.  
XX  
XX 22-MAR-2002; 2002US-0368542P.  
XX  
XX 03-JUN-2002; 2002US-0385362P.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
XX (SUND/) SUNDARACORTHY M.  
XX  
XX (HUDS/) HUDSON B.  
XX  
XX Sundaramoorthy M, Hudson B;  
XX  
XX WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting  
XX basal lamina membrane formation in cell or tissue development.  
XX  
XX Claim 57; SEQ ID NO 289; 168pp; English.  
XX  
XX The present invention describes a crystallised NC1 domain hexamer of type  
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)  
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or  
XX growth; (5) inhibiting endothelial cell interaction with the  
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
XX membrane formation in cell or tissue development; (7) a crystal of an NC1  
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
XX crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,

CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
XX invention.  
XX  
SQ Sequence 20 AA;  
Query Match 8.2%; Score 20; DB 7; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.5e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LQRFMTMPFLFCNVNDVCF 87  
Db 1 LQRFMTMPFLFCNVNDVCF 20  
|||||

RESULT 63  
AAU75606  
ID AAU75606 standard; peptide; 19 AA.  
XX  
AC AAU75606;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Human type IV collagen alpha 3 chain, Tumstatin, mutant T6.  
XX  
XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
XX Tumstatin; angiogenesis; tumour; mutein; mutant.  
XX  
XX Homo sapiens.  
OS  
XX WO200151523-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 08-JAN-2001; 2001WO-US000565.  
XX  
XX 07-JAN-2000; 2000US-00479118.  
XX  
XX 04-APR-2000; 2000US-00543371.  
XX  
XX 21-JUL-2000; 2000US-00625191.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
XX Kalluri R;  
XX  
XX WPI; 2002-188037/24.  
XX  
XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.  
XX  
XX Example 40; Page 133; 205pp; English.  
XX  
XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
XX domain, having one or more of the characteristics selected from: (a) the  
XX ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
XX proliferation of endothelial cells; and (c) the ability to cause  
XX apoptosis of endothelial cells. Also described are the following: (1) use  
XX of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
XX analogue or allelic variant in the preparation of a medicament for  
XX treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
XX where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, T6, which consists of residues 114  
 CC -132 of Tumstatin  
 XX  
 SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 5; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPYISRCTVCEGPATA 131  
 DB 1 RALEPYISRCTVCEGPATA 19

RESULT 64  
 AAU75605  
 ID AAU75605 standard; peptide; 19 AA.

AC AAU75605;  
 DT 08-MAY-2002 (first entry)  
 DE Human type IV collagen alpha 3 chain, Tumstatin, mutant T5.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour; mutein; mutant.

OS Homo sapiens.  
 XX WO200151523-A2.  
 PN 19-JUL-2001.  
 XX 08-JAN-2001; 2001WO-US000565.  
 PF 07-JAN-2000; 2000US-00479118.  
 PR 04-APR-2000; 2000US-00543371.  
 PR 21-JUL-2000; 2000US-00625191.  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;  
 XX WPI; 2002-188037/24.

XX  
 PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.  
 XX

PS Example 40; Page 133; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha5, alpha6, alphav, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, T5, which consists of residues 99-  
 CC 117 of Tumstatin  
 XX

SQ Sequence 19 AA;

Query Match 7.8%; Score 19; DB 5; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALPMNNAPITGRALE 116  
 DB 1 STPALPMNNAPITGRALE 19

RESULT 65  
 ABP58053  
 ID ABP58053 standard; peptide; 19 AA.

XX ABP58053;  
 AC ABP58053;  
 XX 03-MAR-2003 (first entry)

XX Collagen type IV alpha3 chain noncollagenous 1 domain peptide.

DE Angiogenesis; inhibitor; collagen; cytostatic; antiinflammatory;  
 XX immunosuppressive; antiarthritic; antiarteriosclerotic; osteopathic;  
 KW antirheumatic; ophthalmological.

XX Homo sapiens.  
 XX WO200266512-A1.

XX 29-AUG-2002.  
PD 15-FEB-2002; 2002WO-US005211.  
XX 16-FEB-2001; 2001US-0269537P.  
XX 14-SEP-2001; 2001US-0322047P.  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
PA Scialdone MA, Mousa SA, Shuey SW;  
PI WPI; 2003-111767/10.  
XX  
XX  
XX New angiogenesis-inhibitory tripeptide useful for inhibiting endothelial  
PT cell tube formation in angiogenesis-dependent diseases such as cancer,  
PT ocular neovascularization and inflammatory diseases.  
XX  
XX Disclosure; Page 3; 48pp; English.  
XX  
XX The present sequence is that of amino acid residues 185-203 of the  
CC noncollagenous 1 (NC1) domain of the alpha3-chain of basement membrane  
CC collagen type IV. A peptide comprising these residues promoted adhesion  
CC of human melanoma cells by 50-60% over controls and inhibited their  
CC proliferation by 40%. Alanine substitutions through the peptide sequence  
CC indicated that the observed activities were dependent on the presence of  
CC residues 189-191, referred to as the SNS (Ser-Asn-Ser) sequence. The  
CC invention provides methods and compositions for inhibiting endothelial  
CC cell tube formation, the initial step of tumour angiogenesis.  
CC Tripeptides, preferably SNS or SQS (Ser-Gln-Ser) tripeptides, are used to  
CC inhibit angiogenesis-mediated processes such as ocular neovascular  
CC diseases, choroidal neovascular diseases, retina neovascular diseases,  
CC neovascularization of the angle, Bartonellosis, chronic inflammation,  
CC osteoarthritis, rheumatoid arthritis, atherosclerosis phemphigoid,  
CC trachoma, or Osler-Webber-Rendu disease (all claimed). They are also  
CC useful for treating cancer, inflammatory disorders and autoimmune  
CC diseases  
XX  
SQ Sequence 19 AA;  
Query Match 7.8%; Score 19; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 196 CNYYSNSYSFWLASLNP 214  
DB 1 CNYYSNSYSFWLASLNP 19  
RESULT 66  
ADA20269  
ID ADA20269 standard; peptide; 19 AA.  
XX  
XX ADA20269;  
XX  
XX 20-NOV-2003 (first entry)  
DT Human tumstatin deletion peptide T5 amino acid sequence.  
DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytotostatic; gene therapy; alpha 3 chain; tumstatin; human; T5.  
XX  
XX Homo sapiens.  
OS  
XX WO2003059257-A2.  
PN 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
PF 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
PF (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PR 21-DEC-2001; 2001US-00032221.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX Kalluri R;  
XX WPI; 2003-587256/55.  
DR N-PSDB; ADA20224.  
XX  
XX New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
PT  
XX Example 40; Page 131; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins), which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is that of peptide T5, an abridged form of the "tumstatin"  
CC protein of the invention which was derived from the amino acid sequence  
CC of the alpha 3 chain of human type IV collagen.  
XX  
SQ Sequence 19 AA;  
Query Match 7.8%; Score 19; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 98 STPALMPMNMAPITGRALE 116  
DB 1 STPALMPMNMAPITGRALE 19  
RESULT 67  
ADA20265  
ID ADA20265 standard; peptide; 19 AA.  
XX  
XX ADA20265;  
XX  
XX 20-NOV-2003 (first entry)  
DT Human tumstatin deletion peptide T1 amino acid sequence.  
DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
KW metastasis; basement membrane organisation; type IV collagen network;  
KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
KW cytotostatic; gene therapy; alpha 3 chain; tumstatin; human; T1.  
XX  
XX Homo sapiens.  
OS  
XX WO2003059257-A2.  
PN 24-JUL-2003.  
XX  
XX 20-DEC-2002; 2002WO-US040938.  
PF 21-DEC-2001; 2001US-00032221.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PI Kalluri R;  
 XX WPI: 2003-587256/55.  
 DR N-PSDB; ADA20224.  
 XX  
 PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX  
 PS Example 40; Page 131; 240pp; English.  
 XX  
 CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of peptide T1, an abridged form of the "tumstatin"  
 CC protein of the invention which was derived from the amino acid sequence  
 CC of the alpha 3 chain of human type IV collagen.  
 XX  
 XX Sequence 19 AA;

Query Match 7.8%; Score 19; DB 6; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTTRG 19  
 DB 1 GLKGRGDSGSPATWTTRG 19

RESULT 68  
 ADA20270  
 ID ADA20270 standard; peptide; 19 AA.  
 AC  
 XX  
 XX ADA20270;  
 XX  
 XX 20-NOV-2003 (first entry)  
 XX  
 XX Human tumstatin deletion peptide T6 amino acid sequence.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 XX C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 XX cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 XX cytostatic; gene therapy; alpha 3 chain; tumstatin; human; T6.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003059257-A2.  
 XX  
 PD 24-JUL-2003.  
 XX  
 PF 20-DEC-2002; 2002WO-US040938.  
 XX  
 PR 21-DEC-2001; 2001US-00032221.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Kalluri R;  
 XX  
 XX WPI: 2003-587256/55.  
 DR N-PSDB; ADA20224.

XX New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX  
 PS Example 40; Page 131; 240pp; English.  
 XX  
 CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is that of peptide T6, an abridged form of the "tumstatin"  
 CC protein of the invention which was derived from the amino acid sequence  
 CC of the alpha 3 chain of human type IV collagen.  
 XX  
 XX Sequence 19 AA;

Query Match 7.8%; Score 19; DB 6; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPYISRCTVCEGPAIA 131  
 DB 1 RALEPYISRCTVCEGPAIA 19

RESULT 69  
 AAU75601  
 ID AAU75601 standard; peptide; 20 AA.  
 XX  
 AC AAU75601;  
 XX  
 XX 08-MAY-2002 (first entry)  
 XX  
 XX Human type IV collagen alpha 3 chain, Tumstatin, mutant T1.

XX Human; type IV collagen alpha 3 chain; cytostatic; antiangiogenic;  
 XX non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 XX endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 XX Tumstatin; angiogenesis; tumour; mutein; mutant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151523-A2.  
 XX  
 XX 19-JUL-2001.  
 XX  
 XX 08-JAN-2001; 2001WO-US000565.  
 XX  
 PR 07-JAN-2000; 2000US-00479118.  
 PR 04-APR-2000; 2000US-00543371.  
 PR 21-JUL-2000; 2000US-00623191.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Kalluri R;  
 XX  
 XX WPI: 2002-188037/24.  
 XX  
 XX A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 XX treating disorders involving angiogenesis.

Example 40; Page 133; 205pp; English.

PS The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2, or  
 CC alpha3, alpha5, alpha6, alpha7, beta1 or beta2 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC are inhibited by Arresten, Canstatin or Tumstatin; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC of Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 3 chain mutant, T1, which consists of residues 1-  
 CC 20 of Tumstatin

XX Sequence 20 AA;

Query Match 7.8%; Score 19; DB 5; Length 20;

Best Local Similarity 100.0%; Pred. No. 9e-12; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 1 GLXGKRGDSGSPATWTRG 19  
 |||||  
 Db 2 GLXGKRGDSGSPATWTRG 20

RESULT 70  
 ADCL17655

XX ID ADCL17655 standard; peptide; 18 AA.

XX AC ADCL17655;

XX DT 18-DEC-2003 (first entry)

XX Type IV collagen NC1 domain related peptide SEQ ID NO:260.

DE crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
 KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.

OS Synthetic.  
 OS Homo sapiens.

PN WO2003012122-A2.

PD 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX 27-JUL-2001; 2001US-0308523P.

PR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.

PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX (SUND/) SUNDARAMOORTHY M.

XX (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX Claim 57; SEQ ID NO 260; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (6) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (7) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (8) identifying inhibitors of type IV  
 CC domain hexamer of type IV collagen; (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.

XX Sequence 18 AA;

Query Match 7.4%; Score 18; DB 7; Length 18;

Best Local Similarity 100.0%; Pred. No. 8.7e-11;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNDVCFASRND 92

Db 1 PFLFCNVNDVCFASRND 18

RESULT 71

ADCL17672

XX ID ADCL17672 standard; peptide; 18 AA.

XX AC ADCL17672;

XX DT 18-DEC-2003 (first entry)



DE Type IV collagen NC1 domain related peptide SEQ ID NO:277.  
 XX crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO2003012122-A2.  
 XX  
 XX PD 13-FEB-2003.  
 XX  
 XX PF 26-JUL-2002; 2002WO-US023763.  
 XX  
 PR 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX  
 XX Sundaramoorthy M, Hudson B;  
 PI WPI; 2003-332730/31.  
 DR  
 DR New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 XX Claim 57; SEQ ID NO 277; 169pp; English.  
 XX  
 CC The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 18 AA;  
 Query Match 7.4%; Score 18; DB 7; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 210 LNPFRFKPIPTVKAG 227  
 Db 1 LNPFRFKPIPTVKAG 18  
 RESULT 72  
 ADCL17649  
 ID ADC17649 standard; peptide; 18 AA.  
 AC ADC17649.  
 XX  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX  
 XX Type IV collagen NC1 domain related peptide SEQ ID NO:254.  
 DE  
 XX crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO2003012122-A2.  
 XX  
 XX PD 13-FEB-2003.  
 XX  
 XX PF 26-JUL-2002; 2002WO-US023763.  
 XX  
 PR 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX  
 XX Sundaramoorthy M, Hudson B;  
 PI WPI; 2003-332730/31.  
 DR  
 DR New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 XX Claim 57; SEQ ID NO 254; 169pp; English.  
 XX  
 CC The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue

CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
CC invention.

XX Sequence 18 AA;

SQ Query Match

Best Local Similarity 7.4%; Score 18; DB 7; Length 18;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTMPFLFCNVNDVGNFA 88

DB 1 FTMPFLFCNVNDVGNFA 18

RESULT 73

AA118657

ID AA118657 standard; protein; 46 AA.

XX

AC AA118657;

XX

DT 12-OCT-2001 (first entry)

XX

DE Peptide #5091 encoded by probe for measuring cervical gene expression.

XX

KW Probe; human; microarray; gene expression; cervical epithelial cell;

KW Cervical cancer.

XX

OS Homo sapiens.

XX

PN WO200157278-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000670.

XX

PR 04-FEB-2000; 2000US-0180312P.

XX

PR 26-MAY-2000; 2000US-0207456P.

XX

PR 30-JUN-2000; 2000US-00608408.

XX

PR 03-AUG-2000; 2000US-00632366.

XX

PR 21-SEP-2000; 2000US-0234587P.

XX

PR 27-SEP-2000; 2000US-0236359P.

XX

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488901/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX

PS Gene expression in human cervical epithelial cells.

XX

PS Claim 27; SEQ ID NO 23483; 487pp; English.

XX

CC The present invention relates to human single exon nucleic acid probes  
(SENPs: see AA110068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 46 AA;

Query Match

Best Local Similarity 7.0%; Score 17; DB 4; Length 46;

Matches 100.0%; Pred. No. 2.2e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100

DB 10 VCNFASRNDYSYWLSTP 26

RESULT 74

ABG40558

ID ABG40558 standard; peptide; 46 AA.

XX

AC ABG40558;

XX

DT 19-AUG-2002 (first entry)

XX

DE Human peptide encoded by genome-derived single exon probe SEQ ID 30223.

XX

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.

XX Homo sapiens.

OS

PN WO200186003-A2.

XX

PD 15-NOV-2001.

XX

PF 30-JAN-2001; 2001WO-US000665.

XX

PR 04-FEB-2000; 2000US-0180312P.

XX

PR 26-MAY-2000; 2000US-0207456P.

XX

PR 30-JUN-2000; 2000US-00608408.

XX

PR 03-AUG-2000; 2000US-00632366.

XX

PR 21-SEP-2000; 2000US-0234587P.

XX

PR 27-SEP-2000; 2000US-0236359P.

XX

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2002-114183/15.

XX

PT Spatially-addressable set of single exon nucleic acid probes, used to

XX

PS measure gene expression in human lung samples.

XX

PS Claim 27; SEQ ID NO 30223; 634pp; English.

XX

CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC ; the novel set of probes which hybridize at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 46 AA;  
 Query Match 7.0%; Score 17; DB 5; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 VCNFASRNDYSYWLSTP 100  
 DB 10 VCNFASRNDYSYWLSTP 26  
 |||||  
 RESULT 75  
 AAF93524  
 ID AAF93524 standard; protein; 229 AA.  
 XX AC AAF93524;  
 XX DT 25-MAR-2003 (revised)  
 DT 03-OCT-2002 (revised)  
 DT 04-JUN-1990 (first entry)  
 XX Complete sequence of the alpha-1-NC1 domain of type IV collagen.  
 XX Alpha-1-NC1 domain; type IV collagen; cell adhesion; heparin;  
 KW aortic endothelial cells; metastatic carcinoma M4 cells; rat fibroblasts;  
 KW MM fibrosarcoma cells; C6 glioma cell; A431 breast carcinoma cells;  
 KW wound healing; implant acceptance.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT Peptide 17..27  
 FT /note= "TS-3"  
 FT Peptide 49..60  
 FT /note= "TS-2"  
 FT Peptide 201..216  
 FT /note= "TS-1"  
 XX WO8903392-A.  
 XX PD 20-APR-1989.  
 XX PF 20-AUG-1988; 88WO-US003023.  
 XX PR 08-OCT-1987; 87US-00106858.  
 XX PA (MINU ) MINNESOTA UNIVERSITY.  
 XX PI Tailbary EC;  
 XX WPI; 1989-130015/17.  
 XX Polypeptide(s) with type IV collagen activity - used to promote wound

PT healing, implant acceptance and cellular attachment and inhibit malignant  
 PT cells.  
 XX Fig 2; page 1/12; 40pp; English.  
 XX The peptides in the features table are claimed (Claim 1, p. 22). They  
 CC were synthesised using the Merrifield solid phase method. Binding assays  
 CC were carried out using peptides TS-1, TS-2 and TS-3. TS-1 promotes  
 CC adhesions of aortic endothelial cells, metastatic carcinoma M4 cells,  
 CC normal rat fibroblasts, MM fibrosarcoma cells, C6 glioma cells and A431  
 CC breast carcinoma cells. TS-2 binds to type IV collagen, to heparin and  
 CC promotes adhesion of the above cells. Peptides TS-1, TS-2 and TS-3 may be  
 CC used to promote wound healing and implant acceptance, promote cellular  
 CC attachment to culture substrata or inhibit the metastasis of malignant  
 CC cells. They may be used to coat a prosthetic device. (Updated on 03-OCT-  
 CC 2002 to add missing OS field.) (Updated on 25-MAR-2003 to correct PF  
 CC field.) (Updated on 25-MAR-2003 to correct PI field.)  
 XX SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 1; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 VCNFASRNDYSYWLSTP 100  
 DB 70 VCNFASRNDYSYWLSTP 86  
 |||||  
 RESULT 76  
 AAY67943  
 ID AAY67943 standard; protein; 229 AA.  
 XX AC AAY67943;  
 XX DT 03-APR-2000 (first entry)  
 XX Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.  
 XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;  
 KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;  
 KW myocardial angiogenesis disease; Osler-Webber Syndrome; telangiectasia;  
 KW atherosclerosis; scleroderma; plaque neovascularisation; angiofibroma;  
 KW contraception; obesity.  
 XX OS Homo sapiens.  
 XX WO9965940-A1.  
 XX PD 23-DEC-1999.  
 XX PF 17-JUN-1999; 99WO-US013737.  
 XX PR 17-JUN-1998; 98US-0089689P.  
 XX PR 25-MAR-1999; 99US-0126175P.  
 XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX PI Kalluri R;  
 XX WPI; 2000-097708/08.  
 XX N-PSDB; AA257159.

XX Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3  
 PT chain of Type IV collagen used in, e.g. treatment of benign tumors and  
 PT rheumatoid arthritis.  
 XX Example 1; Fig 1B; 117pp; English.  
 XX The present sequence represents the human type IV collagen alpha 1 chain.  
 CC The present invention describes an isolated protein chosen from the NC1  
 CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a

CC fragment, analogue, derivative or mutant, which has anti-angiogenic  
 CC properties. The anti-angiogenic proteins, multimers and chimeras are  
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially  
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign  
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular  
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori  
 CC ulcers, dialysis graft vascular access stenosis, contraception and  
 CC obesity. The compositions can be used to inhibit a disease characterised  
 CC by angiogenic activity, in conjunction with radiation therapy,  
 CC chemotherapy or immunotherapy

XX Sequence 229 AA;

Query Match 7.0%; Score 17; DB 3; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
 |||||  
 Db 70 VCNFASRNDYSYWLSTP 86

RESULT 77

AAU75587

ID AAU75587 standard; protein; 229 AA.

XX AC AAU75587;

DT 08-MAY-2002 (first entry)

DE Human type IV collagen alpha 1 chain.

XX Human; type IV collagen alpha 1 chain; cytostatic; antiangiogenic;  
 KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
 KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
 KW Tumstatin; angiogenesis; tumour.

OS Homo sapiens.

PN WO200151523-A2.

XX 19-JUL-2001.

XX 08-JAN-2001; 2001WO-US000565.

XX 07-JAN-2000; 2000US-00479118.

PR 04-APR-2000; 2000US-00543371.

PR 21-JUL-2000; 2000US-00625191.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2002-188037/24.

DR N-PSDB; ABK15359.

PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
 PT treating disorders involving angiogenesis.

PS Example 1; Fig 1B; 205pp; English.

XX The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1  
 CC domain, having one or more of the characteristics selected from: (a) the  
 CC ability to bind alphavbeta3 integrin; (b) the ability to inhibit  
 CC proliferation of endothelial cells; and (c) the ability to cause  
 CC apoptosis of endothelial cells. Also described are the following: (1) use  
 CC of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue,  
 CC analogue or allelic variant in the preparation of a medicament for  
 CC treating a disorder involving: (a) inhibiting angiogenesis in a tissue,  
 CC where the angiogenesis is mediated by one or more endothelial cell

CC integrins or one or more endothelial cell integrin subunits; or (b) by  
 CC promoting or inducing endothelial cell apoptosis in a tissue, where the  
 CC endothelial cell apoptosis is mediated by one or more endothelial cell  
 CC integrins or one or more endothelial cell integrin subunits; (2) use of  
 CC an antibody or peptide that specifically binds the alpha1, alpha2,  
 CC alpha3, alpha4, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the  
 CC preparation of a medicament for inhibiting angiogenesis or cell  
 CC proliferation; (3) use of an inhibitor, such as an antibody, antibody  
 CC fragment or peptide of receptor-mediated angiogenesis in the preparation  
 CC of a medicament for treating a proliferative disease in a vertebrate,  
 CC where the disease is characterised by angiogenesis that is mediated by  
 CC receptors to Arresten, Canstatin or Tumstatin and where the receptors  
 CC inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one  
 CC or more soluble receptors that bind Arresten, Canstatin or Tumstatin in  
 CC the presence of a medicament for promoting angiogenesis in a tissue; and  
 CC (5) use of integrins in the preparation of a medicament for promoting or  
 CC inducing angiogenesis or cell proliferation in a tissue. The fragments  
 CC Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues  
 CC or allelic variants are useful in the preparation of a medicament for  
 CC treating a disorder involving inhibiting angiogenesis in a tissue, where  
 CC the angiogenesis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits; or by promoting or  
 CC inducing endothelial cell apoptosis in a tissue, where the endothelial  
 CC cell apoptosis is mediated by one or more endothelial cell integrins or  
 CC one or more endothelial cell integrin subunits. The medicament is useful  
 CC in inhibiting tumour growth and for the regression of an established  
 CC tumour. The present sequence represents the amino acid sequence of human  
 CC type IV collagen alpha 1 chain

XX SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 5; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100

|||||

Db 70 VCNFASRNDYSYWLSTP 86

RESULT 78

ADA20217

ID ADA20217 standard; protein; 229 AA.

XX AC ADA20217;

XX 20-NOV-2003 (first entry)

XX Human type IV collagen alpha 1 chain partial protein sequence.

DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; alpha 1 chain; arresten; human.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

DR N-PSDB; ADA20216.

PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX  
XX  
PS Claim 101; Fig 1; 240pp; English.  
XX  
XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumor growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumor  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is the partial amino acid sequence of the alpha 1 chain of human  
CC type IV collagen. The "arresten" peptide of the invention was derived  
CC from this protein.  
XX  
SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 6; Length 229;  
Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 VCNFASRNDYSYWLSTP 100  
DB 70 VCNFASRNDYSYWLSTP 86  
|||||

## RESULT 79

ID ADC17695 7.0%; Score 17; DB 6; Length 229;  
ADCL17699

AC ADC17695;  
XX

DT 18-DEC-2003 (first entry)  
XX

DE Human type IV collagen alpha 1 chain protein SEQ ID NO:302.  
XX

KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumor metastasis inhibitor; tumor growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX

OS Homo sapiens.  
XX

PN WO2003012122-A2.  
XX

PD 13-FEB-2003.  
XX

PF 26-JUL-2002; 2002WO-US023763.  
XX

PR 27-JUL-2001; 2001US-0308523P.  
XX

PR 29-OCT-2001; 2001US-0351289P.  
XX

PR 03-MAR-2002; 2002US-0366854P.  
XX

PR 03-JUN-2002; 2002US-0385362P.  
XX

PA (UNIV ) UNIV KANSAS MEDICAL CENT.  
PA (SUND/) SUNDARAMOORTHY M.  
PA (HUSD/) HUDSON B.

XX Sundaramoorthy M, Hudson B;  
XX  
XX WPI; 2003-332730/31.  
XX

XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.  
XX  
XX Disclosure; SEQ ID NO 302; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
CC IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (1) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents an amino acid sequence which is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 229 AA;

Query Match 7.0%; Score 17; DB 7; Length 229;  
Best Local Similarity 100.0%; Pred. No. 9.2e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 VCNFASRNDYSYWLSTP 100  
DB 70 VCNFASRNDYSYWLSTP 86  
|||||

## RESULT 80

ADCL17699

ID ADC17699 standard; protein; 229 AA.

XX ADC17699;

AC ADC17699;

XX 18-DEC-2003 (first entry)  
XX

DE Human type IV collagen alpha 5 chain protein SEQ ID NO:306.  
XX

KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumor metastasis inhibitor; tumor growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX

OS Homo sapiens.  
XX

PN WO2003012122-A2.  
XX

PD 13-FEB-2003.  
 XX 26-JUL-2002; 2002WO-US023763.  
 XX 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND// SUNDARAMOORTHY M.  
 PA (HUDS// HUDSON B.  
 XX Sundaramoorthy M, Hudson B;  
 XX WPI; 2003-332730/31.  
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 XX Disclosure; SEQ ID NO 306; 168pp; English.  
 XX  
 XX The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and  
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents an amino acid sequence which is used in the exemplification of  
 CC the present invention.  
 XX  
 XX Sequence 229 AA;  
 XX  
 XX Query Match 7.0%; Score 17; DB 7; Length 229;  
 XX Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
 XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 84 VCNFASRNDYSYWLSTP 100  
 DB 70 VCNFASRNDYSYWLSTP 86  
 |||||  
 RESULT 81  
 ID AAY31991  
 ID AAY31991 standard; protein; 260 AA.  
 XX AAY31991;  
 AC AAY31991;  
 XX  
 XX 05-JAN-2000 (first entry)  
 XX  
 XX Type IV collagen NC1 domain alpha-1 monomer.  
 DE Type IV collagen; NC1 domain; non-collagenous domain; human;  
 XX angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
 KW rheumatoid arthritis; retinal neovascularization;  
 KW

KW choroidal neovascularization; macular degeneration;  
 KW corneal neovascularization; retinopathy of prematurity;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidemic keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; sogrens; acne rosacea; phlyctenulosis;  
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;  
 KW ulcer; herpes simplex infection; Herpes zoster infection;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anaemia;  
 KW sarcoid; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;  
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;  
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu;  
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;  
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;  
 KW pemphigoid.  
 XX  
 XX Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Location/Qualifiers  
 XX Key  
 XX Peptide 1..17 /note= "BM40 signal peptide"  
 XX Protein 18..260 /note= "mature protein"  
 XX Peptide 18..25 /note= "affinity tag"  
 XX Protein 26..260 /note= "NCI alpha-1 monomer"  
 XX  
 XX WC9949885-A2.  
 XX  
 XX 07-OCT-1999.  
 XX  
 XX 26-MAR-1999; 99WO-US006445.  
 XX  
 XX 27-MAR-1998; 98US-0079783P.  
 XX 29-OCT-1998; 98US-0106170P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 XX  
 XX Hudson BG, Sarrae MP;  
 XX  
 XX WPI; 1999-501297/51.  
 XX N-PSDB; AA220089.  
 XX  
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer  
 XX useful for treating e.g. tumor growth or metastasis, neovascularisation,  
 XX etc.  
 XX  
 XX Disclosure; Fig 17a; 56pp; English.  
 XX  
 XX This sequence represents a recombinant type IV collagen non-collagenous  
 XX (NC1) domain alpha-1 polypeptide composed of a BM40 signal sequence  
 XX (which is cleaved from the mature protein) to facilitate protein  
 XX secretion, and a mature protein comprising an affinity tag (facilitates  
 XX purification and identification of the material) and the alpha-1 chain  
 XX monomer. The invention provides methods and kits for inhibiting  
 XX angiogenesis, tumour growth and metastasis, and endothelial cell  
 XX interaction with the extracellular matrix, each method comprising  
 XX contacting the tumour or animal tissue with 1 or more isolated type IV  
 XX collagen NC1 alpha chain monomer(s) selected from the group consisting of  
 XX alpha-1, alpha-2, alpha-3 and alpha-6 NCI chain monomers (see AAY31991-  
 XX 96). The monomers can be produced via recombinant protein expression. The  
 XX polynucleotides and polypeptides are used to treat an angiogenesis  
 XX mediated disorder or condition, especially selected from solid and blood-  
 XX borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal  
 XX neovascularization, choroidal neovascularization, macular degeneration,

CC corneal neovascularization, retinopathy of prematurity, corneal graft  
CC rejection, neovascular glaucoma, retrolental fibroplasia, epidemic  
CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic  
CC keratitis, superior limbic keratitis, pterygium keratitis sicca, seborrheic  
CC acne rosacea, phlyctenulosis, syphilis, mycobacterial infections, lipid  
CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes  
CC simplex infections, herpes zoster infections, protozoan infections, Kaposi's  
CC sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal  
CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's  
CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,  
CC sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,  
CC vein occlusion, artery occlusion, carotid obstructive disease, chronic  
CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets  
CC disease, myopia, optic pits, Stargards disease, pars planitis, chronic  
CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser  
CC complications, abnormal proliferation of fibrovascular tissue,  
CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,  
CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative  
CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)  
XX Sequence 260 AA:

This sequence is a human type IV collagen alpha chain monomer, designated alpha(IV)NC1. The invention relates to a method for inhibiting angiogenesis, tumour growth or metastases, or endothelial cell interactions with the extracellular matrix, comprising contacting the cells or tissue with a polypeptide composition containing antagonists of specific integrin receptors. The methods and the antagonists are useful for inhibiting angiogenesis, tumour growth or metastases, or endothelial cell interaction with the extracellular matrix. The antagonists are also useful for treating diseases and conditions with accompanying undesired angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas, carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma, neuroblastoma, osteosarcoma or leukaemia). These are also applicable to treating non-tumorigenic diseases and conditions with accompanying undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis, retinal neovascularisation, choroidal neovascularisation, muscular degeneration, corneal graft rejection, vitamin A deficiency, atopic keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma, sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser complications, chronic inflammation or psoriasis

Sequence 260 AA:



```

FT Peptide /note= "mature protein"
FT 18..25
FT /note= "affinity tag"
FT 26..264
FT Protein /note= "NC1 alpha-5 monomer"
FT
PN WO9949885-A2.
XX 07-OCT-1999.
XX
XX 26-MAR-1999; 99WO-US006445.
XX
XX 27-MAR-1998; 98US-0079783P.
XX 29-OCT-1998; 98US-0106170P.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG, Sarraz MP;
XX
XX WPI; 1999-601297/51.
XX N-PSDB; AAZ20093.
XX
XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
XX useful for treating e.g. tumor growth or metastasis, neovascularisation,
XX etc.
XX
XX Disclosure; Fig 17e; 56pp; English.
XX
XX This sequence represents a recombinant type IV collagen non-collagenous
XX (NC1) domain alpha-5 polypeptide composed of a BV40 signal sequence
XX (which is cleaved from the mature protein) to facilitate protein
XX secretion, and a mature protein comprising an affinity tag (facilitates
XX purification and identification of the material) and the alpha-1 chain
XX monomer. The invention provides methods and kits for inhibiting
XX angiogenesis, tumour growth and metastasis, and endothelial cell
XX interaction with the extracellular matrix, each method comprising
XX contacting the tumour or animal tissue with 1 or more isolated type IV
XX collagen NC1 alpha chain monomer(s) selected from the group consisting of
XX alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see A431991-
XX 96). The monomers can be produced via recombinant protein expression. The
XX polynucleotides and polypeptides are used to treat an angiogenesis-
XX mediated disorder or condition, especially selected from solid and blood-
XX borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal
XX neovascularization, choroidal neovascularization, macular degeneration,
XX corneal neovascularization, retinopathy of prematurity, corneal graft
XX rejection, neovascular glaucoma, retrolental fibroplasia, epidemic
XX keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic
XX acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid
XX degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes
XX simplex infections, herpes zoster infections, protozoan infections,
XX Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal
XX keratolysis, trauma, systemic lupus, polyarteritis, Wegener's
XX sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,
XX sickle cell anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease,
XX vein occlusion, artery occlusion, carotid obstructive disease, chronic
XX uveitis, chronic vitritis, Lyme's disease, Bales disease, Behcets
XX disease, myopia, optic pits, Sargarts disease, pars planitis, chronic
XX retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser
XX complications, abnormal proliferation of fibrovascular tissue,
XX haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,
XX osteoarthritis, chronic inflammation, Crohn's disease, ulcerative
XX colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)
XX
XX SQ Sequence 264 AA;
XX
XX Query Match 7.0%; Score 17; DB 2; Length 264;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 84 VCNFASRNDYSYWLSTP 100
XX |||||
XX 105 VCNFASRNDYSYWLSTP 121
XX
XX DB

```

## RESULT 84

AA97557

ID AA97557 standard; protein; 264 AA.

XX

AC AA97557;

XX

DT 12-FEB-2001 (first entry)

XX

DE Human alpha5(IV)NC1 protein sequence.

XX

XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;  
 KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;  
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;  
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;  
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;  
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;  
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;  
 KW chronic inflammation; psoriasis; therapy; alpha5(IV)NC1.

XX

OS Homo sapiens.

XX

PN WO200059532-A1.

XX

PD 12-OCT-2000.

XX

XX 31-MAR-2000; 2000WO-US008678.

XX

XX 01-APR-1999; 99US-0127391P.

XX

XX (BIOS-) BIOSTRATUM INC.

XX

XX Brooks P, Hudson B;

XX

XX WPI; 2000-664962/64.

XX

XX N-PSDB; AAA90995.

XX

XX Use of antagonists of specific integrin receptors for inhibiting  
 PT angiogenesis, tumor growth or metastases, or endothelial cell  
 PT interactions with the extracellular matrix.

XX

XX Disclosure; Fig 17e; 78pp; English.

XX

XX This sequence is a human type IV collagen alpha chain monomer, designated  
 CC alpha5(IV)NC1. The invention relates to a method for inhibiting  
 CC angiogenesis, tumor growth or metastases, or endothelial cell  
 CC interactions with the extracellular matrix, comprising contacting the  
 CC cells or tissue with a polypeptide composition containing antagonists of  
 CC specific integrin receptors. The methods and the antagonists are useful  
 CC for inhibiting angiogenesis, tumor growth or metastases, or endothelial  
 CC cell interaction with the extracellular matrix. The antagonists are also  
 CC useful for treating diseases and conditions with accompanying undesired  
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,  
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,  
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to  
 CC treating non-tumorigenic diseases and conditions with accompanying  
 CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,  
 CC retinal neovascularisation, choroidal neovascularisation, muscular  
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic  
 CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,  
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser  
 CC complications, chronic inflammation or psoriasis

XX

SQ Sequence 264 AA;

XX

Query Match 7.0%; Score 17; DB 3; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;

XX

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

QY 84 VCNFASRNDYSYWLSTP 100

XX

105 VCNFASRNDYSYWLSTP 121

XX

DB

```
RESULT 85
AAB54044
ID AAB54044 standard; protein; 309 AA.
XX AC AAB54044;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:496.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
XX KW diagnosis; identification; cytostatic; neuroprotective; nootropic;
XX KW immunomodulatory; relaxant; contraceptive; gynaecological;
XX KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX KW linkage analysis; tissue identification; tissue typing; forensic; neural;
XX KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
XX KW cardiovascular; renal; proliferative.
XX OS Homo sapiens.
XX PN WO200055320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005989.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-579444/54.
XX DR N-PSDB; AAC98809.
XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX PT treating, or ameliorating a medical condition, particular pancreatic
XX PT cancer, or for use in assays for diagnosing a pathological condition.
XX PS Claim 11; Page 934-935; 1379pp; English.
XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX CC proteins, called pancreatic cancer antigens, given in AAB54008 to
XX CC AAB54466. The human pancreatic cancer antigens have cytostatic,
XX CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
XX CC gynaecological, cardiant and antiinflammatory activities, and can be used
XX CC in gene therapy. The polynucleotide and proteins can be used for
XX CC preventing, treating, or ameliorating a medical condition or in assays
XX CC for diagnosing a pathological condition or a susceptibility to one in a
XX CC subject. Binding partners to the proteins and the activity of the
XX CC proteins can be identified. The pancreatic cancer antigens can be used to
XX CC detect, treat or prevent pancreatic disorders, especially cancer.
XX CC Agonists and antagonists to the antigens can be screened for. The
XX CC pancreatic cancer antigen polynucleotides can be used to design nucleic
XX CC acid hybridisation probes that can be used in chromosome mapping, linkage
XX CC analysis, tissue identification and/or typing and a variety of forensic
XX CC and diagnostic methods. The proteins can be used to generate antibodies
XX CC which are used to purify, detect and target the polypeptides, including
XX CC both in vivo and in vitro diagnostic and therapeutic methods. The
XX CC proteins can be used to treat or prevent neural, immune system, muscular,
XX CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX CC sequences used in the exemplification of the present invention
XX SQ Sequence 309 AA;
Query Match 7.0%; Score 17; DB 3; Length 309;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100

RESULT 86
AAB58169
ID AAB58169 standard; protein; 406 AA.
XX AC AAB58169;
XX DT 14-MAR-2001 (first entry)
XX DE Lung cancer associated polypeptide sequence SEQ ID 507.
XX KW Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnerary;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease.
XX OS Homo sapiens.
XX PN WO200055180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005918.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR WPI: 2000-587514/55.
XX DR N-PSDB; AAF18045.
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer.
XX PS Claim 11; Page 996-998; 1425pp; English.
XX CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58549. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective, cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the protein
XX CC or polynucleotide sequences. The lung cancer associated polynucleotide
XX CC sequences may be used for detection of lung cancer, chromosome
XX CC identification, as chromosome markers, and for numerous other diagnostic
XX CC or research purposes. The proteins may be used to treat disorders such as
XX CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
XX CC cardiovascular, renal, and proliferative disorders. The proteins may also
XX CC be used in the treatment of wounds and infectious diseases.
XX CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
XX CC used in the course of the invention for the identification and
XX CC characterisation of the polynucleotide and protein sequences
XX SQ Sequence 406 AA;
Query Match 7.0%; Score 17; DB 3; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
Db 247 VCNFASRNDYSYWLSTP 263
```

RESULT 88  
AAW09643



AC ABB57334;  
XX  
XX  
XX 07-MAR-2002 (first entry)  
XX  
XX Mouse ischaemic condition related protein sequence SEQ ID NO:933.  
DE  
XX  
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.  
KW  
XX  
XX Mus musculus.  
XX  
XX WO200188188-A2.  
PN  
XX  
XX 22-NOV-2001.  
PD  
XX  
XX 18-MAY-2001; 2001WO-JP004192.  
PF  
XX  
XX 18-MAY-2000; 2000JP-00145977.  
PR  
XX  
XX (UYN1-) UNIV NIHOON SCHOOL JURIDICAL PERSON.  
PA  
XX  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
PI  
XX  
XX WPI; 2002-034733/04.  
DR  
XX  
XX N-PSDB; ABI99819.  
DR  
XX  
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or by  
PT determining the expression profile of a gene group comprising these  
PT genes.  
XX  
XX Claim 2; Page 2352-2359; 2690pp; English.  
PS  
XX  
XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (ABI99202 to ABI9912, encoding the  
CC protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischaemic condition-improving drugs or  
CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR  
CC primers for a mouse ischaemic condition related sequence, which are used  
CC in the exemplification of the present invention  
XX  
XX Sequence 1669 AA;  
SQ  
Query Match 7.0%; Score 17; DB 5; Length 1669;  
Best Local Similarity 100.0%; Pred. No. 5.6e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 84 VCNFASRNDYSYWLSTP 100  
DB 1510 VCNFASRNDYSYWLSTP 1526  
RESULT 92  
ABU54467  
ID ABU54467 standard; protein; 1669 AA.  
XX  
XX ABU54467;  
AC  
XX  
XX 12-MAR-2003 (first entry)  
DT  
XX  
XX Human tumour endothelial marker TEM 31.  
DE  
XX  
XX Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;  
KW Tumour endothelial marker; normal endothelial marker; PEM;  
KW pan-endothelial marker; polycystic kidney disease; psoriasis;  
KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;  
KW

KW neangiogenesis; immune response; cytostatic; antidiabetic;  
KW ophthalmological; anti-rheumatic; antiarthritic; antipsoriatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200283874-A2.  
PN  
XX  
XX 24-OCT-2002.  
PD  
XX  
XX 10-APR-2002; 2002WO-US008253.  
PF  
XX  
XX 11-APR-2001; 2001US-0282850P.  
PR  
XX  
XX 06-FEB-2002; 2002US-0354262P.  
PR  
XX  
XX (UYJO) UNIV JOHNS HOPKINS.  
PA  
XX  
XX Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;  
PI  
XX  
XX WPI; 2003-093016/08.  
DR  
XX  
XX N-PSDB; ABX72039.  
DR  
XX  
XX New purified human transmembrane protein, designated as tumor endothelial  
PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors, or  
PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or  
PT psoriasis.  
XX  
XX Disclosure; Page 272-275; 374pp; English.  
PS  
XX  
XX The present invention relates to a novel method for the isolation of  
CC endothelial cells (ECs), and the identification of genes expressed in  
CC normal and tumour ECs. Tumour endothelial marker (TEM), normal  
CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are  
CC identified in human ECs. The human EC marker proteins and the  
CC polynucleotide sequences encoding them are useful for detecting,  
CC diagnosing or treating tumors as well as polycystic kidney disease,  
CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also  
CC useful for inhibiting neangiogenesis or tumour angiogenesis, for  
CC inducing an immune response to tumour endothelial cells in a patient, or  
CC for identifying candidate drugs for treating tumors. The present  
CC sequence represents a human TEM or NEM protein of the invention  
XX  
XX Sequence 1669 AA;  
SQ  
Query Match 7.0%; Score 17; DB 6; Length 1669;  
Best Local Similarity 100.0%; Pred. No. 5.6e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 84 VCNFASRNDYSYWLSTP 100  
DB 1510 VCNFASRNDYSYWLSTP 1526  
RESULT 93  
AAM39077  
ID AAM39077 standard; protein; 1672 AA.  
XX  
XX AAM39077;  
AC  
XX  
XX 22-OCT-2001 (first entry)  
DT  
XX  
XX Human polypeptide SEQ ID NO 2222.  
DE  
XX  
XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200153312-A1.  
PN  
XX

```

PD 26-JUL-2001.
XX
PF
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
XX
PR 21-JAN-2000; 2000US-00488725.
XX
PR 25-APR-2000; 2000US-00552317.
XX
PR 20-JUN-2000; 2000US-00598042.
XX
PR 19-JUL-2000; 2000US-00620312.
XX
PR 03-AUG-2000; 2000US-00653450.
XX
PR 14-SEP-2000; 2000US-00662191.
XX
PR 19-OCT-2000; 2000US-00693036.
XX
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac R;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AA158233.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2222; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 1672 AA;
SQ
Query Match 7.0%; Score 17; DB 4; Length 1672;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
DB 1513 VCNFASRNDYSYWLSTP 1529
RESULT 94
ABG04839
ID ABG04839 standard; protein; 1685 AA.
XX
AC ABG04839;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #4830.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
PD 26-JUL-2001.
XX
PF
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
XX
PR 21-JAN-2000; 2000US-00488725.
XX
PR 25-APR-2000; 2000US-00552317.
XX
PR 20-JUN-2000; 2000US-00598042.
XX
PR 19-JUL-2000; 2000US-00620312.
XX
PR 03-AUG-2000; 2000US-00653450.
XX
PR 14-SEP-2000; 2000US-00662191.
XX
PR 19-OCT-2000; 2000US-00693036.
XX
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac R;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AA158233.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2222; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 1672 AA;
SQ
Query Match 7.0%; Score 17; DB 4; Length 1685;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 VCNFASRNDYSYWLSTP 100
DB 1526 VCNFASRNDYSYWLSTP 1542
RESULT 95
ABG15619
ID ABG15619 standard; protein; 1693 AA.
XX
AC ABG15619;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #15610.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
PF

```

```
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS79806.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 45978; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1693 AA;
XX
XX Query Match 7.0%; Score 17; DB 4; Length 1693;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 84 VCNFASRNDXYWLTSTP 100
XX | | | | | | | | | | | | | | | | | |
XX 1534 VCNFASRNDXYWLTSTP 1550
XX
XX RESULT 96
XX ADC17470
XX ID ADC17470 standard; peptide; 16 AA.
XX AC ADC17470;
XX AD 18-DEC-2003 (first entry)
XX DE Type IV collagen NCI domain related peptide SEQ ID NO:74.
XX
XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;
XX antianemic; ophthalmological; antiarteriosclerotic; antiulcer;
XX endothelial cell adhesion inhibitor;
XX endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
XX ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
XX blood-borne tumour.
XX Synthetic.
XX
XX OS
XX Homo sapiens.
XX WO2003012122-A2.
XX 13-FEB-2003.
XX 26-JUL-2002; 2002WO-US023763.
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX (SUND/) SUNDARAMOORTHY M.
XX (HUDS/) HUDSON B.
XX Sundaramoorthy M, Hudson B;
XX WPI; 2003-332730/31.
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT basal lamina membrane formation in cell or tissue development.
XX Disclosure; SEQ ID NO 74; 169pp; English.
XX The present invention describes a crystallised NCI domain hexamer of type
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC growth; (6) inhibiting endothelial cell interaction with the
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC membrane formation in cell or tissue development; (7) a crystal of type IV
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC antipsoriatic, antianemic, ophthalmological, antiarteriosclerotic and
CC antiulcer activities, and can be used as an inhibitor of angiogenesis
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC used for treating an angiogenesis-mediated disease or condition
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC or for inhibiting basal lamina membrane formation in cell or tissue
CC development. The methods are useful for inhibiting angiogenesis in
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC cell interaction with the extracellular matrix in an animal tissue, and
CC identifying inhibitors of type IV collagen assembly. The present sequence
CC represents a peptide which is used in the exemplification of the present
CC invention.
XX
XX Sequence 16 AA;
XX
XX Query Match 6.6%; Score 16; DB 7; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 8.8e-09;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 50 VQGNQRAHGQDLGTLG 65
XX | | | | | | | | | | | | | | | | | |
XX 1 VQGNQRAHGQDLGTLG 16
XX
XX RESULT 97
XX AAE09491
XX ID AAE09491 standard; peptide; 15 AA.
XX AC AAE09491;
XX 19-NOV-2001 (first entry)
XX Human C2 alpha-3 peptide to construct alpha/alpha3 (IV)NC1 protein.
XX
```



XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;  
KW Goodpasture syndrome; C2 alpha-3 peptide.  
XX  
OS Homo sapiens.  
XX  
PN US6277558-B1.  
XX  
PD 21-AUG-2001.  
XX  
PF 12-NOV-1999; 99US-00439897.  
XX  
PR 30-NOV-1990; 90US-00621091.  
PR 07-MAR-1995; 95US-00399889.  
PR 07-OCT-1998; 98US-00167364.  
XX  
PA (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
PI Hudson BG;  
XX  
DR WPI; 2001-540401/60.  
XX  
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting  
PT Goodpasture antibodies from bodily fluid/tissue from patient or for  
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with  
PT the polypeptide.  
XX  
PS Claim 1; Fig 12; 46pp; English.  
XX  
CC The invention relates to a method for detecting Goodpasture antibodies  
CC from a bodily fluid or tissue of a patient. The method comprises  
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)  
CC collagen polypeptide that contains a conformational epitope for the  
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for  
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a  
CC patient, and for treating Goodpasture syndrome in a patient. The present  
CC sequence is human alpha chain peptide used for constructing human  
CC alpha/alpha3 (IV) NC1 fusion protein  
XX  
SQ Sequence 15 AA;  
XX  
Query Match 6.1%; Score 15; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 29 TAIPSCPEGVPLYS 43  
DB |||||||  
1 TAIPSCPEGVPLYS 15  
XX  
RESULT 98  
AAE09497  
ID AAE09497 standard; peptide; 15 AA.  
XX  
AC AAE09497;  
XX  
DT 19-NOV-2001 (first entry)  
XX  
DE Human C5 alpha-3 peptide to construct alpha/alpha3 (IV) NC1 protein.  
XX  
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;  
KW Goodpasture syndrome; C5 alpha-3 peptide.  
XX  
OS Homo sapiens.  
XX  
PN US6277558-B1.  
XX  
PD 21-AUG-2001.  
XX  
PF 12-NOV-1999; 99US-00439897.  
XX  
PR 30-NOV-1990; 90US-00621091.  
PR 07-MAR-1995; 95US-00399889.  
XX  
PS Claim 1; Fig 12; 46pp; English.

PR 07-OCT-1998; 98US-00167364.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
PI Hudson BG;  
XX  
DR WPI; 2001-540401/60.  
XX  
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting  
PT Goodpasture antibodies from bodily fluid/tissue from patient or for  
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with  
PT the polypeptide.  
XX  
PS Example 19; Fig 12; 46pp; English.  
XX  
CC The invention relates to a method for detecting Goodpasture antibodies  
CC from a bodily fluid or tissue of a patient. The method comprises  
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)  
CC collagen polypeptide that contains a conformational epitope for the  
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for  
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a  
CC patient, and for treating Goodpasture syndrome in a patient. The present  
CC sequence is human alpha chain peptide used for constructing human  
CC alpha/alpha3 (IV) NC1 fusion protein  
XX  
SQ Sequence 15 AA;  
XX  
Query Match 6.1%; Score 15; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 102 LMPNMNMAPITGRALE 116  
DB |||||||  
1 LMPNMNMAPITGRALE 15  
XX  
RESULT 99  
AAE09499  
ID AAE09499 standard; peptide; 15 AA.  
XX  
AC AAE09499;  
XX  
DT 19-NOV-2001 (first entry)  
XX  
DE Human C6 alpha-3 peptide to construct alpha/alpha3 (IV) NC1 protein.  
XX  
XX Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;  
KW Goodpasture syndrome; C6 alpha-3 peptide.  
XX  
OS Homo sapiens.  
XX  
PN US6277558-B1  
XX  
PD 21-AUG-2001.  
XX  
PF 12-NOV-1999; 99US-00439897.  
XX  
PR 30-NOV-1990; 90US-00621091.  
PR 07-MAR-1995; 95US-00399889.  
PR 07-OCT-1998; 98US-00167364.  
XX  
PA (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
PI Hudson BG;  
XX  
DR WPI; 2001-540401/60.  
XX  
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting  
PT Goodpasture antibodies from bodily fluid/tissue from patient or for  
PT treating Goodpasture syndrome by contacting bodily fluid or tissue with  
PT the polypeptide.  
XX  
PS Claim 1; Fig 12; 46pp; English.



CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 7; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLNPFMRKPIPT 223  
 DB 1 SLNPFMRKPIPT 15  
 |||||

## RESULT 102

ADCL17586  
 ID ADCL17586 standard; peptide; 15 AA.

XX AC ADCL17586;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:191.

XX KW crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.

XX PA (SUND/) SUNDARAMOORTHY M.

XX PA (HUSD/) HUDSON B.

XX PI Sundaramoorthy M, Hudson B;

XX DR WPI; 2003-332730/31.

XX PT New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.

XX PS Disclosure; SEQ ID NO 191; 168pp; English.

XX CC The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)

CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
 CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 15 AA;

Query Match 6.1%; Score 15; DB 7; Length 15;

Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 FRKPIPTVKAGELE 230  
 DB 1 FRKPIPTVKAGELE 15  
 |||||

## RESULT 103

ADCL17607

ID ADCL17607 standard; peptide; 15 AA.

XX AC ADCL17607;

XX DT 18-DEC-2003 (first entry)

XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:212.

XX KW crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.

XX OS Synthetic.

OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX PA (UNIV ) UNIV KANSAS MEDICAL CENT.

XX PA (SUND/) SUNDARAMOORTHY M.

XX PA (HUSD/) HUDSON B.

PI Sundaramoorthy M, Hudson B;  
 XX WPI; 2003-332730/31.  
 XX  
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 XX Disclosure; SEQ ID NO 212; 168pp; English.  
 PS  
 XX The present invention describes a crystallised NCI domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (6) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (7) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and  
 CC antitumor activities, and can be used as an inhibitor of angiogenesis,  
 CC tumor growth, tumor metastasis, rheumatoid arthritis or blood-borne tumors  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 15 AA;  
 SQ  
 Query Match 6.1%; Score 15; DB 7; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 DVCNFASGRNDYSYWL 97  
 Db 1 DVCNFASGRNDYSYWL 15  
 RESULT 104  
 ADCL17626  
 ID ADCL17626 standard; peptide; 15 AA.  
 XX  
 AC ADCL17626;  
 XX  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX Type IV collagen NCI domain related peptide SEQ ID NO:231.  
 DE  
 XX crystallised NCI domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
 KW antianemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO2003012122-A2.  
 PN  
 XX

PD 13-FEB-2003.  
 XX  
 PF 26-JUL-2002; 2002WO-US023763.  
 XX  
 PR 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND)/ SUNDARAMOORTHY M.  
 PA (HUDS)/ HUDSON B.  
 XX  
 PI Sundaramoorthy M, Hudson B;  
 DR WPI; 2003-332730/31.  
 XX  
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 XX Disclosure; SEQ ID NO 231; 168pp; English.  
 PS  
 XX The present invention describes a crystallised NCI domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (6) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (7) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antianemic, ophthalmological, antiarteriosclerotic and  
 CC antitumor activities, and can be used as an inhibitor of angiogenesis,  
 CC tumor growth, tumor metastasis, rheumatoid arthritis or blood-borne tumors  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 15 AA;  
 SQ  
 Query Match 6.1%; Score 15; DB 7; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 193 RGTCTNYNSYSYFWL 207  
 Db 1 RGTCTNYNSYSYFWL 15  
 RESULT 105  
 ADCL17532  
 ID ADCL17532 standard; peptide; 15 AA.  
 XX  
 AC ADCL17532;  
 XX  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX Type IV collagen NCI domain related peptide SEQ ID NO:137.  
 DE  
 XX crystallised NCI domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW



CC invention.  
XX Sequence 15 AA;  
SQ

Query Match 6.1%; Score 15; DB 7; Length 15;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 TSAGSEGTGQALASP 175  
Db 1 TSAGSEGTGQALASP 15

RESULT 107  
ADCL17560  
ID ADCL17560 standard; peptide; 14 AA.  
AC ADCL17560;  
XX  
XX  
XX 18-DEC-2003 (first entry)  
XX  
XX Type IV collagen NC1 domain related peptide SEQ ID NO:165.  
XX  
XX crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX WO2003012122-A2.  
PN  
XX 13-FEB-2003.  
PD  
XX 26-JUL-2002; 2002WO-US023763.  
PF  
XX 27-JUL-2001; 2001US-0308523P.  
PR  
XX 29-OCT-2001; 2001US-0351289P.  
PR  
XX 22-MAR-2002; 2002US-0366854P.  
PR  
XX 03-JUN-2002; 2002US-0385362P.  
PR  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
PA (SUND/) SUNDARAMOORTHY M.  
PA (HUSD/) HUDSON B.  
XX Sundaramoorthy M, Hudson B;  
PI WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.  
XX  
XX Disclosure; SEQ ID NO 165; 168pp; English.  
XX  
XX The present invention describes a crystallised NC1 domain hexamer of type  
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,

CC antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and  
CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis, tumours  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
XX invention.  
SQ Sequence 14 AA;

Query Match 5.7%; Score 14; DB 7; Length 14;  
Best Local Similarity 100.0%; Pred. No. 8.8e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 158 IMFTSAGSEGTGQA 171  
Db 1 IMFTSAGSEGTGQA 14

RESULT 108  
ADCL17427  
ID ADCL17427 standard; peptide; 14 AA.  
XX  
XX AC ADCL17427;  
XX  
XX 18-DEC-2003 (first entry)  
DT  
XX  
XX Type IV collagen NC1 domain related peptide SEQ ID NO:28.  
XX  
XX crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX WO2003012122-A2.  
PN  
XX 13-FEB-2003.  
PD  
XX 26-JUL-2002; 2002WO-US023763.  
PF  
XX 27-JUL-2001; 2001US-0308523P.  
PR  
XX 29-OCT-2001; 2001US-0351289P.  
PR  
XX 22-MAR-2002; 2002US-0366854P.  
PR  
XX 03-JUN-2002; 2002US-0385362P.  
PR  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
PA (SUND/) SUNDARAMOORTHY M.  
PA (HUSD/) HUDSON B.  
XX Sundaramoorthy M, Hudson B;  
PI WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.  
XX

PS Disclosure; SEQ ID NO 28; 169pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
CC antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and  
CC antitumor activities, and can be used as an inhibitor of angiogenesis,  
CC tumor growth, tumor metastasis, rheumatoid arthritis or blood-borne tumors  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumor metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
CC invention.

XX Sequence 14 AA;  
SQ

Query Match 5.7%; Score 14; DB 7; Length 14;  
Best Local Similarity 100.0%; Pred. No. 8.8e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 SCLQRTTMTPLFC 79  
DB 1 SCLQRTTMTPLFC 14  
|||||

RESULT 109  
ADCI17508  
ID ADCI17508 standard; peptide; 14 AA.  
XX  
AC ADCI17508;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Type IV collagen NC1 domain related peptide SEQ ID NO:112.  
XX  
KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumor metastasis inhibitor; tumor growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.

OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO2003012122-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 26-JUL-2002; 2002WO-US023763.  
XX  
PR 27-JUL-2001; 2001US-0308523P.  
PR 29-OCT-2001; 2001US-0351289P.  
PR 22-MAR-2002; 2002US-036854P.  
PR 03-JUN-2002; 2002US-0385362P.

(UNIV ) UNIV KANSAS MEDICAL CENT.  
(SUND/) SUNDARAMOORTHY M.  
(HUDS/) HUDSON B.  
XX Sundaramoorthy M, Hudson B;  
XX WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
PT basal lamina membrane formation in cell or tissue development.  
XX  
XX Disclosure; SEQ ID NO 112; 169pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
CC growth; (5) inhibiting endothelial cell interaction with the  
CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
CC antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and  
CC antitumor activities, and can be used as an inhibitor of angiogenesis,  
CC tumor growth, tumor metastasis, endothelial cell adhesion, endothelial  
CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
CC used for treating an angiogenesis-mediated disease or condition  
CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors  
CC or for inhibiting basal lamina membrane formation in cell or tissue  
CC development. The methods are useful for inhibiting angiogenesis in  
CC tissue, inhibiting tumor metastasis or growth, inhibiting endothelial  
CC cell interaction with the extracellular matrix in an animal tissue, and  
CC identifying inhibitors of type IV collagen assembly. The present sequence  
CC represents a peptide which is used in the exemplification of the present  
CC invention.

XX Sequence 14 AA;  
SQ

Query Match 5.7%; Score 14; DB 7; Length 14;  
Best Local Similarity 100.0%; Pred. No. 8.8e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 SFLFVQGNQRAHQ 59  
DB 1 SFLFVQGNQRAHQ 14  
|||||

RESULT 110  
ADCI17402  
ID ADCI17402 standard; peptide; 14 AA.  
XX  
AC ADCI17402;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Type IV collagen NC1 domain related peptide SEQ ID NO:3.  
XX  
KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumor metastasis inhibitor; tumor growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.



```

OS Synthetic.
OS Homo sapiens.
PN WO2003012122-A2.
PD 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 5; SEQ ID NO 3; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX antitumor activities, and can be used as an inhibitor of angiogenesis
XX tumor growth, tumor metastasis, endothelial cell adhesion, endothelial
XX cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 14 AA;
XX
XX Query Match 5.7%; Score 14; DB 7; Length 14;
XX Best Local Similarity 100.0%; Pred. No. 8.8e-07;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 75 PFLFCNVNDVCNFA 88
XX
XX Db 1 PFLFCNVNDVCNFA 14
XX
XX RESULT 111
XX ADC17667
XX
XX ID ADC17667 standard; peptide; 14 AA.
XX
XX AC ADC17667;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX

```

Type IV collagen NCI domain related peptide SEQ ID NO:272.

crystallised NCI domain hexamer of type IV collagen; angiogenesis inhibitor; angiogenesis-mediated disease; tumour metastasis inhibitor; tumour growth inhibitor; endothelial cell interaction inhibitor; basal lamina membrane formation inhibitor; cytostatic; antiproliferative; antianaemic; ophthalmological; antiarteriosclerotic; antitumor; endothelial cell adhesion inhibitor; ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis; blood-borne tumour.

Synthetic.

Homo sapiens.

WO2003012122-A2.

13-FEB-2003.

26-JUL-2002; 2002WO-US023763.

27-JUL-2001; 2001US-0308523P.

29-OCT-2001; 2001US-0351289P.

22-MAR-2002; 2002US-0366854P.

03-JUN-2002; 2002US-0385362P.

(UNIV ) UNIV KANSAS MEDICAL CENT.

(SUND/) SUNDARAMOORTHY M.

(HUDS/) HUDSON B.

Sundaramoorthy M, Hudson B;

WPI; 2003-332730/31.

New polypeptide, useful for treating an angiogenesis-mediated disease or condition consisting of glaucoma or blood-borne tumors or for inhibiting basal lamina membrane formation in cell or tissue development.

Claim 5; SEQ ID NO 3; 168pp; English.

The present invention describes a crystallised NCI domain hexamer of type IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a pharmaceutical composition comprising the polypeptide and a carrier; (3) inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated disease or condition in a mammal; (5) inhibiting tumour metastasis or growth; (5) inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue; (6) inhibiting basal lamina membrane formation in cell or tissue development; (7) a crystal of an NCI domain hexamer of type IV collagen; (8) identifying inhibitors of type IV collagen assembly; and (9) an inhibitor of type IV collagen assembly. A crystallised NCI domain hexamer of type IV collagen (I) has cytostatic, antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and antitumor activities, and can be used as an inhibitor of angiogenesis tumor growth, tumor metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly. A (I) polypeptide can be used for treating an angiogenesis-mediated disease or condition consisting of glaucoma, sickle cell anaemia, ulcerative colitis, psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors or for inhibiting basal lamina membrane formation in cell or tissue development. The methods are useful for inhibiting angiogenesis in tissue, inhibiting tumour metastasis or growth, inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue, and identifying inhibitors of type IV collagen assembly. The present sequence represents a peptide which is used in the exemplification of the present invention.

Sequence 14 AA;

Query Match 5.7%; Score 14; DB 7; Length 14; Best Local Similarity 100.0%; Pred. No. 8.8e-07; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

75 PFLFCNVNDVCNFA 88

1 PFLFCNVNDVCNFA 14

RESULT 111

ADC17667

ADC17667 standard; peptide; 14 AA.

ADC17667;

18-DEC-2003 (first entry)

Qy 186 PFLECHGRCNY 199  
 Db 1 PFLECHGRCNY 14

RESULT 112  
 ADC17605  
 ID ADC17605 standard; peptide; 15 AA.  
 XX  
 AC ADC17605;  
 XX  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 XX  
 DE Type IV collagen NC1 domain related peptide SEQ ID NO:210.  
 XX  
 KW crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytoskeletal; antipsoriatic;  
 KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX  
 FN WO2003012122-A2.  
 XX  
 XX  
 PD 13-FEB-2003.  
 XX  
 XX  
 PF 26-JUL-2002; 2002WO-US023763.  
 XX  
 PR 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX  
 PI Sundaramoorthy M, Hudson B;  
 XX  
 XX  
 DR WPI; 2003-332730/31.  
 XX  
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 PS Disclosure; SEQ ID NO 210; 168pp; English.  
 XX

The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytoskeletal,  
 CC antipsoriatic, anti-anaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumors  
 CC or for inhibiting basal lamina membrane formation in cell or tissue

CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 5.7%; Score 14; DB 7; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 84 VCNFASRNDYSYWL 97  
 Db 2 VCNFASRNDYSYWL 15  
 RESULT 113  
 AAB03336  
 ID AAB03336 standard; peptide; 16 AA.  
 XX  
 AC AAB03336;  
 XX  
 DT 08-SEP-2000 (first entry)  
 XX  
 XX  
 DE Human epitope P23.  
 XX  
 KW Human; rhesus blood group system; Rh; Rhd; Rhce; sickle cell disease;  
 KW thalassaemia; Rhc; Rhc; Rhe; alloimmunisation prevention;  
 KW autoimmune Rh haemolytic disease; rhesus protein; immunosuppressive;  
 KW vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200032632-A2.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 01-DEC-1999; 99WO-GB004027.  
 XX  
 PR 01-DEC-1998; 98GB-00026378.  
 XX  
 PA (UYAB-) UNIV ABERDEEN.  
 PA (COMM-) COMMON SERVICES AGENCY SCOTTISH HEALTH S.  
 XX  
 PI Urbaniak SJ, Barker RN;  
 XX  
 DR WPI; 2000-412291/35.  
 XX  
 PT Composition for prevention of alloimmunization or immunosuppression of a  
 PT response elicited by alloimmunization or an autoimmune hemolytic disease,  
 PT comprises an epitope of a rhesus protein.  
 XX  
 PS Disclosure; Page 82; 92pp; English.  
 XX  
 CC Human blood contains the rhesus (Rh) blood group system, and humans can  
 CC either be Rhd positive or negative. This can lead to complications during  
 CC transfusions or pregnancy if Rhd negative individuals are exposed to Rhd  
 CC positive blood, leading to them becoming immunised to produce anti-D. The  
 CC present invention relates to new human allo- and auto-reactive T-cell  
 CC epitopes (AA99760-Y99769 and AAB03201-B03337) from Rhd, Rhc, Rhe, Rhd  
 CC and Rhd proteins. These epitopes bind to T-cells to elicit an immune  
 CC response, i.e. immunisation. These epitopes can be used as a vaccine for  
 CC the prevention of alloimmunisation or immunosuppression of a response  
 CC elicited by alloimmunisation or an autoimmune haemolytic disease.  
 CC Examples of autoimmune haemolytic diseases are sickle cell disease and  
 CC thalassaemia  
 XX  
 SQ Sequence 16 AA;  
 Query Match 5.7%; Score 14; DB 3; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 KIISRCQVCMKKRH 244  
DB 3 KIISRCQVCMKKRH 16

RESULT 114  
AA95913  
ID AA95913 standard; peptide; 21 AA.

XX AC AA95913;  
XX AC  
XX DT 20-NOV-2000 (first entry)  
XX DE Human Goodpasture antigen N-terminal peptide GPpepl1a9.  
XX KW Goodpasture antigen binding protein; GPBP; GPpepl; human;  
XX KW autoimmune disease; apoptosis; cancer; tumour; diagnosis; therapy;  
XX KW mutant; mutein.  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX FN WO2000050607-A2.  
XX PD 31-AUG-2000.  
XX PF 24-FEB-2000; 2000WO-IB000324.  
XX FR 24-FEB-1999; 99US-0121483P.  
XX PA (SAUS/) SAUS J.  
XX PI Saus J;  
XX DR WPI; 2000-572094/53.  
XX DT Novel Goodpasture antigen binding proteins useful for diagnosing and  
XX FT treating autoimmune disorders, tumor, and preventing cell apoptosis.  
XX PS Example 1; Page 21; 15pp; English.

XX CC The present sequence is that of GPpepl1a9, comprising the N-terminal 21  
XX CC amino acids of human Goodpasture antigen (GP) but carrying a Ser-9 to Ala  
XX CC amino acid substitution. The peptide was used to characterise the  
XX CC phosphorylation activity of human Goodpasture binding protein (GPBP, see  
XX CC AA95913), a novel serine/threonine kinase that specifically binds to and  
XX CC phosphorylates native GPpepl. The invention provides nucleic acids (see  
XX CC AA50341-53) encoding GPBP recombinant vectors, host cells, encoded  
XX CC polypeptides (see AA95900-11) and antibodies. It also provides methods  
XX CC for detecting the presence of an autoimmune condition or apoptosis by  
XX CC detecting an increase in GPBP expression, and methods for treating an  
XX CC autoimmune disorder, apoptosis or a tumour by modifying GPBP expression  
XX CC or activity, especially using a GP-derived peptide

XX SQ Sequence 21 AA;  
Query Match 5.7%; Score 14; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTTRGFVFT 23  
DB 8 GSPATWTTTRGFVFT 21

RESULT 115  
ABG79203  
ID ABG79203 standard; peptide; 21 AA.  
XX AC ABG79203;  
XX AC

DT 15-NOV-2002 (first entry)  
XX DE Human Goodpasture protein peptide, GPpepl1a9 mutant.  
XX KW Goodpasture antigen binding protein; Goodpasture syndrome; antigen;  
XX KW chromosome 5q13; neuroprotective; dermatological; immunosuppressive;  
XX KW autoimmune condition; phosphorylation; myelin basic protein; MBP;  
XX KW alpha3 type IV collagen non-collagenous domain; NCI; multiple sclerosis;  
XX KW systemic lupus erythematosus; cutaneous lupus erythematosus; pemphigus;  
XX KW pemphigoid; lichen planus; human.  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX PN WO200261430-A2.  
XX PD 08-AUG-2002.  
XX PF 31-JAN-2002; 2002WO-EP001010.  
XX PR 31-JAN-2001; 2001US-0265249P.  
XX PA (SAUS/) SAUS J.  
XX PI Saus J;  
XX DR WPI; 2002-619280/66.  
XX PT Identifying candidate compounds for treating autoimmune conditions, e.g.  
XX PT Goodpasture syndrome or lupus, comprises identifying compounds that  
XX PT reduce phosphorylation of, or formation of conformational isomers of,  
XX PT target proteins.

XX PS Example 1; Page 26; 217pp; English.

XX CC The invention relates to identifying candidate compounds to treat an  
XX CC autoimmune condition by identifying compounds that reduce phosphorylation  
XX CC of a first target protein (I) (which is selected from Goodpasture antigen  
XX CC binding protein (GPBP), an alpha3 type IV collagen non-collagenous (NCI)  
XX CC domain polypeptide comprising Lys-Gly-Lys-Arg-Gly-Asp-Ser-Gly-Ser-Pro-  
XX CC Ala-Thr-Trp-Thr-Arg-Gly-Phe-Val-Phe-Thr, and a polypeptide comprising  
XX CC Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly), or reduce formation of  
XX CC conformational isomers of the second target protein (II) (selected from  
XX CC an alpha3 type IV collagen NCI domain polypeptide and myelin basic  
XX CC protein, MBP). Also included are (1) an isolated type IV collagen alpha3  
XX CC NCI domain conformational isomer, which has an amino acid sequence  
XX CC identical to the wild type alpha3 type IV collagen NCI domain, is  
XX CC stabilised by disulphide bonds, and has a molecular weight in a non-  
XX CC reducing sodium dodecyl sulphate gel of 22, 23, 25, 27, or 28 kD, and in  
XX CC a reducing sodium dodecyl sulphate gel of 29 kDa; and (2) an isolated  
XX CC type IV collagen alpha3 NCI domain. The human gene for GPBP is located on  
XX CC chromosome 5q13. The method is useful for treating autoimmune conditions,  
XX CC such as Goodpasture Syndrome, multiple sclerosis, systemic and cutaneous  
XX CC lupus erythematosus, pemphigus, pemphigoid and lichen planus. The present  
XX CC sequence represents an alpha3 type IV collagen non-collagenous (NCI)  
XX CC domain (also known as the GP antigen) peptide antigen

XX SQ Sequence 21 AA;  
Query Match 5.7%; Score 14; DB 5; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTTRGFVFT 23  
DB 8 GSPATWTTTRGFVFT 21

RESULT 116  
ADA20237  
ID ADA20237 standard; peptide; 25 AA.  
XX AC ADA20237;  
XX AC

XX 20-NOV-2003 (first entry)

XX T7 mutant peptide related to human type IV collagen and angiogenesis.

XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;

XX metastasis; basement membrane organisation; type IV collagen network;

XX C-terminal globular non-collagenous domain; NC1; type IV collagen;

XX cell surface receptor; integrin; angiogenic activity; protein synthesis;

XX cytostatic; gene therapy; T7 mutant peptide; mutant; mutagen;

XX type IV collagen alpha 3 chain; tumstatin; human.

XX Synthetic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 5 /note= "Wild-type Leu substituted by Met"

XX Misc-difference 9 /note= "Wild-type Val substituted by Ile"

XX Misc-difference 11 /note= "Wild-type Asp substituted by Asn"

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002/ 2002WO-US040938.

XX 21-DEC-2001/ 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI; 2003-587256/55.

XX New peptide, useful for preparing a composition for inhibiting tumor

XX growth, angiogenic activity or protein synthesis in a mammalian tissue.

XX Claim 61; Page 45; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments

XX with anti-angiogenic properties. The invention also relates to the DNA

XX sequences which encode the novel proteins. A wide variety of diseases are

XX the result of undesirable angiogenesis. The formation of new capillaries

XX from pre-existing vessels is essential for tumour growth and metastasis.

XX Basement membrane organisation is dependent on the assembly of a type IV

XX collagen network which may occur through the C-terminal globular non-

XX collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2

XX forms are ubiquitously exhibited in human basement membranes. In the

XX present invention, cell surface receptors (in particular integrins) which

XX specifically bind anti-angiogenic proteins and peptides (in particular

XX the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV

XX collagen) are disclosed. The proteins of the invention may inhibit tumour

XX growth, angiogenic activity in mammalian tissue or protein synthesis in

XX endothelial cells and thus may exhibit cytostatic activity. The DNA

XX sequences of the invention may be useful in gene therapy. The present

XX invention is the amino acid sequence of the mutated T7 peptide of the

XX sequence. The wild-type T7 peptide sequence is given in Seq ID37 (see

XX ADA20236) and this was derived from the amino acid sequence of tumstatin,

XX which in turn was derived from the amino acid sequence of human type IV

XX collagen alpha 3 chain.

XX Sequence 25 AA;

XX Query Match 5.7%; Score 14; DB 6; Length 25;

XX Best Local Similarity 100.0%; Pred. No. 1.5e-06;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWL 97

DB 12 VCNFASRNDYSYWL 25

RESULT 117

AA44173

ID AAY44173 standard; peptide; 12 AA.

XX AAY44173;

XX 01-FEB-2000 (first entry)

XX Bovine type IV collagen alpha3 chain protein epitope motif.

XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;

XX Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.

XX Bos taurus

XX US5973120-A.

XX 26-OCT-1999.

XX 07-MAR-1995; 95US-00399889.

XX 30-NOV-1990; 90US-00621091.

XX (UYVA ) UNIV YALE.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Reenders ST, Morrison KE;

XX WPI; 1999-610317/52.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for

XX diagnosis and treatment of Goodpasture syndrome.

XX Claim 3; Col 36; 27pp; English.

XX This sequence represents an epitope from the bovine alpha3 chain of type

XX IV collagen polypeptide (AAY44171). The collagen polypeptide chain is

XX useful for detecting Goodpasture antibodies in blood or tissue from a

XX human patient and for treating Goodpasture syndrome especially by

XX neutralising the antibodies in the blood. The polypeptides also have a

XX nephrotrophic activity

XX Sequence 12 AA;

XX Query Match 4.9%; Score 12; DB 2; Length 12;

XX Best Local Similarity 100.0%; Pred. No. 8.6e-05;

XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISRCQVCVKKQH 244

DB 1 ISRCQVCVKKQH 12

RESULT 118

AA56785

ID AAY56785 standard; peptide; 12 AA.

XX AAY56785;

XX 27-MAR-2000 (first entry)

XX Human alpha3 type IV collagen C-terminal domain fragment.

XX Goodpasture syndrome; type IV collagen; alpha3 chain; human.

XX Homo sapiens.

XX US6007980-A.

XX 28-DEC-1999.

```

XX CC The invention relates to a method for detecting Goodpasture antibodies
CC from a bodily fluid or tissue of a patient. The method comprises
CC contacting the bodily fluid or tissue with alpha-3 chain type (IV)
CC collagen polypeptide that contains a conformational epitope for the
CC Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
CC detecting Goodpasture antibodies from a bodily fluid or tissue from a
CC patient, and for treating Goodpasture syndrome in a patient. The present
CC sequence is human alpha-3 chain of type IV collagen peptide
XX
SQ Sequence 12 AA;
Query Match 4.9%; Score 12; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISRCQVCMKKRH 244
DB 1 ISRCQVCMKKRH 12
RESULT 120
AAE09493
ID AAE09493 standard; peptide; 12 AA.
AC AAE09493;
AD
DT 19-NOV-2001 (first entry)
DX
DE Human C3 alpha-3 peptide to construct alpha1/alpha3(IV)NC1 protein.
DX
KW Human; alpha-3 chain; type IV collagen; immunosuppressive; therapy;
KW Goodpasture syndrome; C3 alpha-3 peptide.
XX
XX Homo sapiens.
XX
XX US6277558-B1.
XX
XX 21-AUG-2001.
XX
XX 12-NOV-1999; 99US-00439897.
XX
XX 30-NOV-1990; 90US-00621091.
XX
XX 07-MAR-1995; 95US-00399889.
XX
XX 07-OCT-1998; 98US-00167364.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX Hudson BG;
XX
XX WFI; 2001-540401/60.
XX
XX Use of alpha (IV) noncollagenous 1 collagen polypeptide for detecting
XX Goodpasture antibodies from bodily fluid/tissue from patient or for
XX treating Goodpasture syndrome by contacting bodily fluid or tissue with
XX the polypeptide.
XX
XX Example 19; Fig 12; 46pp; English.
XX
XX The invention relates to a method for detecting Goodpasture antibodies
XX from a bodily fluid or tissue of a patient. The method comprises
XX contacting the bodily fluid or tissue with alpha-3 chain type (IV)
XX collagen polypeptide that contains a conformational epitope for the
XX Goodpasture antibodies. Alpha-3 chain of type (IV) collagen is useful for
XX detecting Goodpasture antibodies from a bodily fluid or tissue from a
XX patient, and for treating Goodpasture syndrome in a patient. The present
XX sequence is human alpha chain peptide used for constructing human
XX alpha1/alpha3 (IV) NC1 fusion protein
XX
SQ Sequence 12 AA;
Query Match 4.9%; Score 12; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;

```



CC mediated disease condition, where the antigen specific antibody is  
CC produced by an antigen-reactive B cell population present in a host. The  
CC toxin is useful for treating allergies, viral diseases conditions, and  
CC autoimmune disorders. Also treated are skin diseases; autoimmune  
CC endocrinopathies; vasculitic syndromes; cardiovascular disease;  
CC immunohaematologic disorders; gastrointestinal diseases; neurologic  
CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;  
CC and infertility disorders. The present sequence represents a collagen IV  
CC alpha 3 domain epitope peptide. An antibody response to this antigen is  
CC implicated in Goodpasture syndrome, a disorder which may be treated using  
CC the toxin of the invention  
XX  
SQ Sequence 12 AA;

Query Match 4.9%; Score 12; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 155 PSFIMFTSAGSE 166  
|||||  
Db 1 PSFIMFTSAGSE 12

RESULT 123  
AAB97333  
ID AAB97333 standard; peptide; 12 AA.  
AC AAB97333;  
XX  
DT 13-AUG-2001 (first entry)  
DE Collagen IV alpha 3 domain epitope peptide #1.  
XX  
KW B cell; toxin; antigen specific; antibody mediated disease; virucide;  
KW immunosuppressive; antiinflammatory; anti-allergic; antidiabetic;  
KW thyromimetic; antithyroid; vasotropic; cardiac; antiulcer;  
KW neuroprotective; antirheumatic; antiarthritic; dermatological;  
KW ophthalmological; nephrotropic; allergy; autoimmune disorder;  
KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;  
KW cardiovascular disease; immunohaematologic disorder; neurologic disease;  
KW gastrointestinal disease; collagen vascular disease; renal diseases;  
KW pulmonary disease; infertility disorder; collagen IV;  
KW Goodpasture syndrome.

XX Unidentified.  
XX WO200132853-A1.  
XX 10-MAY-2001.  
XX 12-OCT-2000; 2000WO-US028157.  
XX 29-OCT-1999; 99US-0162464P.  
XX (BIOM-) INST APPLIED BIOMEDICINE.  
XX Chaplin JW;  
XX WPI; 2001-316435/33.  
XX B cell clonal toxin useful for treating autoimmune disorders such as  
PT Grave's disease, myocardial infarction, Crohn's disease, multiple  
PT sclerosis, comprises a group that causes toxin to be internalized by B  
PT cell.  
XX Disclosure; Page 35; 46pp; English.

XX This invention relates to a B cell clonal toxin. The toxin is made from  
XX two moieties, the first causes the toxin to be internalised by a B cell,  
CC and the second is a biologically acceptable toxin. The invention includes  
CC a method for inactivating/killing an antigen specific B cell. A target B  
CC cell is contacted with an effective amount of a B cell clonal toxin. The  
CC method is useful for selective immunosuppression in conditions

CC characterised by the presence of an unwanted or deleterious immune  
CC response, e.g. in the treatment of antigen specific antibody mediated  
CC disease conditions. Use of the B cell clonal toxin can result in  
CC immunosuppressive; antiinflammatory; anti-allergic; virucide; antidiabetic  
CC ; thyromimetic; antithyroid; vasotropic; cardiac; antiulcer;  
CC neuroprotective; antirheumatic; antiarthritic; dermatological;  
CC ophthalmological; and nephrotropic activity. The toxin is particularly  
CC useful for treating a host suffering from an antigen specific antibody  
CC mediated disease condition, where the antigen specific antibody is  
CC produced by an antigen-reactive B cell population present in a host. The  
CC toxin is useful for treating allergies, viral disease conditions, and  
CC autoimmune disorders. Also treated are skin diseases; autoimmune  
CC endocrinopathies; vasculitic syndromes; gastrointestinal diseases; neurologic  
CC immunohaematologic disorders; gastrointestinal diseases; pulmonary diseases;  
CC and infertility disorders. The present sequence represents a collagen IV  
CC alpha 3 domain epitope peptide. An antibody response to this antigen is  
CC implicated in Goodpasture syndrome, a disorder which may be treated using  
CC the toxin of the invention  
XX  
SQ Sequence 12 AA;

Query Match 4.9%; Score 12; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 SGSPATWTTRGF 20  
|||||  
Db 1 SGSPATWTTRGF 12

RESULT 124  
ABP58056  
ID ABP58056 standard; peptide; 12 AA.  
XX  
AC ABP58056;  
XX  
DT 03-MAR-2003 (first entry)  
DE Peptide used in endothelial cell tube formation inhibition assay.  
XX  
KW Angiogenesis; inhibitor; collagen; cytostatic; antiinflammatory;  
KW immunosuppressive; antiarthritic; antiarteriosclerotic; osteopathic;  
KW antirheumatic; ophthalmological.  
XX  
OS Synthetic.

XX Key Location/Qualifiers  
XX Modified-site 1 /note= "N-terminal acetylation"  
XX  
XX WO200266512-A1.  
XX 29-AUG-2002.  
XX 15-FEB-2002; 2002WO-US005211.  
XX 16-FEB-2001; 2001US-0269537P.  
XX 14-SEP-2001; 2001US-0322047P.  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX Scialdone MA, Mousa SA, Shuey SW;  
XX WPI; 2003-111767/10.

XX New angiogenesis-inhibitory tripeptide useful for inhibiting endothelial  
PT cell tube formation in angiogenesis-dependent diseases such as cancer,  
PT ocular neovascularization and inflammatory diseases.  
XX  
XX Example 1; Page 27; 48pp; English.  
XX  
XX The present sequence is that of an acetylated peptide including a Ser-Asn



-Ser tripeptide sequence. The peptide was prepared using a standard solid phase synthesis protocol using Fmoc chemistry. Inhibition of human fibroblast growth factor basic (FGF2)-stimulated human umbilical vein endothelial cells (HUVEC) by the peptide was determined in an example from the invention. Inhibition of EC tube formation by the peptide (0.015  $\mu$ M) was 55 +/- 2% (area) and 49 +/- 1% (length). Corresponding values for the tripeptide acetyl-Ser-Asn-Ser-carboxamide were 97 +/- 8% and 84 +/- 9%, respectively, showing the tripeptide to be a more potent inhibitor of angiogenesis than the larger peptide. The invention provides methods and compositions for inhibiting endothelial cell tube formation, the initial step of tumour angiogenesis. Tripeptides, preferably SNS (Ser-Asn-Ser) or SOS (Ser-Gln-Ser), are used to inhibit angiogenesis-mediated processes such as ocular neovascular diseases, choroidal neovascular diseases, retina neovascular diseases, neovascularization of the angle, Bartorellitis, chronic inflammation, osteoarthritis, rheumatoid arthritis, atherosclerosis phlegmipoid, trachoma, or Osler-Weber-Rendu disease (all claimed). They are also useful for treating cancer, inflammatory disorders and autoimmune diseases

Sequence 12 AA;

Query Match 4.9%; Score 12; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 CNYYSNSYSFWL 207  
|||||||  
DB 1 CNYYSNSYSFWL 12

RESULT 125

ADCL17409  
ID ADCL17409 standard; peptide; 12 AA.

AC ADCL17409;

DT 18-DEC-2003 (first entry)

DE Type IV collagen NC1 domain related peptide SEQ ID NO:10.

KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

PR 27-JUL-2001; 2001US-0308523P.

PR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.

PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

PA (SUND/) SUNDARAMOORTHY M.

PA (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX

PT New polypeptide, useful for treating an angiogenesis-mediated disease or condition consisting of glaucoma or blood-borne tumours or for inhibiting basal lamina membrane formation in cell or tissue development.

PS Claim 10; SEQ ID NO 10; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a pharmaceutical composition comprising the polypeptide and a carrier; (3) inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated disease or condition in a mammal; (5) inhibiting tumour metastasis or growth; (5) inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue; (6) inhibiting basal lamina membrane formation in cell or tissue development; (7) a crystal of an NC1 domain hexamer of type IV collagen; (8) identifying inhibitors of type IV collagen assembly; and (9) an inhibitor of type IV collagen assembly. A crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic, antipsoriatic, antianaemic, ophthalmological, antiarteriosclerotic and antiulcer activities, and can be used as an inhibitor of angiogenesis, tumour growth, tumour metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly. A (I) polypeptide can be used for treating an angiogenesis-mediated disease or condition consisting of glaucoma, sickle cell anaemia, ulcerative colitis, psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours or for inhibiting basal lamina membrane formation in cell or tissue development. The methods are useful for inhibiting angiogenesis in tissue, inhibiting tumour metastasis or growth, inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue, and identifying inhibitors of type IV collagen assembly. The present sequence represents a peptide which is used in the exemplification of the present invention.

XX Sequence 12 AA;

Query Match 4.9%; Score 12; DB 7; Length 12;

Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 PFLECHGRGTCN 197  
|||||||  
DB 1 PFLECHGRGTCN 12

RESULT 126

ADCL17634

ID ADCL17634 standard; peptide; 11 AA.

AC ADCL17634;

XX 18-DEC-2003 (first entry)

XX Type IV collagen NC1 domain related peptide SEQ ID NO:239.

XX crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.

XX Synthetic.

OS Homo sapiens.

XX WO2003012122-A2.

XX 13-FEB-2003.

XX 26-JUL-2002; 2002WO-US023763.

XX

```

PR 27-JUL-2001; 2001US-0308523P.
PR 29-OCT-2001; 2001US-0351289P.
PR 22-MAR-2002; 2002US-0366854P.
PR 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 55; SEQ ID NO 239; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (1) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX antitumour activities, and can be used as an inhibitor of angiogenesis,
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 11 AA;
XX
XX Query Match 4.5%; Score 11; DB 7; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.00085;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 210 LNPERMFRKPI 220
Db |||||
1 LNPERMFRKPI 11
XX
RESULT 127
ADCI17571
ID ADCI17571 standard; peptide; 11 AA.
XX
XX ADCI17571;
XX
XX 18-DEC-2003 (first entry)
XX
XX Type IV collagen NCI domain related peptide SEQ ID NO:176.
XX
XX crystallised NCI domain hexamer of type IV collagen;
XX angiogenesis inhibitor; angiogenesis-mediated disease;
XX tumour metastasis inhibitor; tumour growth inhibitor;
XX endothelial cell interaction inhibitor;
XX basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
XX antianaemic; ophthalmological; antiarteriosclerotic; antitumour;
XX endothelial cell adhesion inhibitor;

```

---

```

KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW blood-borne tumour.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX WO2003012122-A2.
XX
XX 13-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-US023763.
XX
XX 27-JUL-2001; 2001US-0308523P.
XX 29-OCT-2001; 2001US-0351289P.
XX 22-MAR-2002; 2002US-0366854P.
XX 03-JUN-2002; 2002US-0385362P.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
PA (SUND/) SUNDARAMOORTHY M.
PA (HUDS/) HUDSON B.
XX
XX Sundaramoorthy M, Hudson B;
XX
XX WPI; 2003-332730/31.
XX
XX New polypeptide, useful for treating an angiogenesis-mediated disease or
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX basal lamina membrane formation in cell or tissue development.
XX
XX Claim 43; SEQ ID NO 176; 168pp; English.
XX
XX The present invention describes a crystallised NCI domain hexamer of type
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX growth; (5) inhibiting endothelial cell interaction with the
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX membrane formation in cell or tissue development; (7) a crystal of an NCI
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
XX crystallised NCI domain hexamer of type IV collagen (1) has cytostatic,
XX antiproliferative, antianaemic, ophthalmological, antiarteriosclerotic and
XX antitumour activities, and can be used as an inhibitor of angiogenesis,
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be
XX used for treating an angiogenesis-mediated disease or condition
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX or for inhibiting basal lamina membrane formation in cell or tissue
XX development. The methods are useful for inhibiting angiogenesis in
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX cell interaction with the extracellular matrix in an animal tissue, and
XX identifying inhibitors of type IV collagen assembly. The present sequence
XX represents a peptide which is used in the exemplification of the present
XX invention.
XX
XX Sequence 11 AA;
XX
XX Query Match 4.5%; Score 11; DB 7; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.00085;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 189 ECHGRGTCNY 199
Db |||||
1 ECHGRGTCNY 11
XX
RESULT 128
ADCI17545
ID ADCI17545 standard; peptide; 11 AA.
XX
XX

```

AC ADCL17545;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Type IV collagen NC1 domain related peptide SEQ ID NO:150.  
XX  
KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiprosclerotic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
XX WO2003012122-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 26-JUL-2002; 2002WO-US023763.  
XX  
XX 27-JUL-2001; 2001US-0308523P.  
XX 29-OCT-2001; 2001US-0351289P.  
XX 22-MAR-2002; 2002US-0366854P.  
XX 03-JUN-2002; 2002US-0385362P.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX (SUND/) SUNDARAMOORTHY M.  
XX (HUDS/) HUDSON B.  
XX Sundaramoorthy M, Hudson B;  
XX WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting  
XX basal lamina membrane formation in cell or tissue development.  
XX  
XX Claim 36; SEQ ID NO 150; 168pp; English.  
XX  
XX The present invention describes a crystallised NC1 domain hexamer of type  
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)  
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or  
XX growth; (6) inhibiting endothelial cell interaction with the  
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
XX membrane formation in cell or tissue development; (7) a crystal of an NC1  
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
XX crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
XX antiprosclerotic, antianaemic, ophthalmological, antiarteriosclerotic and  
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,  
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
XX cell interaction with the extracellular matrix in an animal tissue, and  
XX identifying inhibitors of type IV collagen assembly. The present sequence  
XX represents a peptide which is used in the exemplification of the present  
XX invention.  
XX  
XX Sequence 11 AA;  
SQ

Query Match 4.5%; Score 11; DB 7; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00085;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 102 LMPNMAPITG 112  
DB 1 LMPNMAPITG 11  
RESULT 129  
ADCL17666  
ID ADCL17666 standard; peptide; 14 AA.  
XX  
XX AC ADCL17666;  
XX DT 18-DEC-2003 (first entry)  
XX  
XX DE Type IV collagen NC1 domain related peptide SEQ ID NO:271.  
XX  
XX KW crystallised NC1 domain hexamer of type IV collagen;  
KW angiogenesis inhibitor; angiogenesis-mediated disease;  
KW tumour metastasis inhibitor; tumour growth inhibitor;  
KW endothelial cell interaction inhibitor;  
KW basal lamina membrane formation inhibitor; cytostatic; antiprosclerotic;  
KW antianaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
KW endothelial cell adhesion inhibitor;  
KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
KW blood-borne tumour.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
XX WO2003012122-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 26-JUL-2002; 2002WO-US023763.  
XX  
XX 27-JUL-2001; 2001US-0308523P.  
XX 29-OCT-2001; 2001US-0351289P.  
XX 22-MAR-2002; 2002US-0366854P.  
XX 03-JUN-2002; 2002US-0385362P.  
XX  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX (SUND/) SUNDARAMOORTHY M.  
XX (HUDS/) HUDSON B.  
XX Sundaramoorthy M, Hudson B;  
XX WPI; 2003-332730/31.  
XX  
XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
XX condition consisting of glaucoma or blood-borne tumors or for inhibiting  
XX basal lamina membrane formation in cell or tissue development.  
XX  
XX Claim 57; SEQ ID NO 271; 168pp; English.  
XX  
XX The present invention describes a crystallised NC1 domain hexamer of type  
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)  
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or  
XX growth; (6) inhibiting endothelial cell interaction with the  
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
XX membrane formation in cell or tissue development; (7) a crystal of an NC1  
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
XX crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
XX antiprosclerotic, antianaemic, ophthalmological, antiarteriosclerotic and  
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,  
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be

CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 14 AA;  
 SQ

Query Match 4.5%; Score 11; DB 7; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 ECHGRGTCNY 199  
 Db 4 ECHGRGTCNY 14  
 |||||

RESULT 130  
 ADA20240  
 ID ADA20240 standard; peptide; 19 AA.  
 AC ADA20240;  
 XX  
 XX 20-NOV-2003 (first entry)  
 DT  
 XX TP3 peptide related to human type IV collagen alpha and angiogenesis.  
 DE anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 XX metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytosolic; gene therapy; TP3 peptide; tumstatin; human;  
 KW type IV collagen alpha 3 chain; mutant; mutein.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 1 /note= "Wild-type Phe substituted by Lys"  
 FT Misc-difference 8 /note= "Wild-type Asp substituted by Cys"  
 FT  
 XX WO2003059257-A2.  
 PN  
 XX 24-JUL-2003.  
 PD  
 XX 20-DEC-2002; 2002WO-US040938.  
 XX  
 XX 21-DEC-2001; 2001US-00032221.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA  
 XX Kalluri R;  
 PI  
 XX WPI; 2003-587256/55.  
 DR  
 XX New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 PT  
 XX Claim 64; Page 45; 240pp; English.  
 PS  
 XX This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV

CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the TP3 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.  
 XX  
 XX Sequence 19 AA;  
 SQ

Query Match 4.5%; Score 11; DB 6; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 0.0014;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYS 94  
 Db 9 VCNFASRNDYS 19  
 |||||

RESULT 131  
 ADC17413  
 ID ADC17413 standard; peptide; 22 AA.  
 XX  
 XX ADC17413;  
 AC  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX Type IV collagen NC1 domain related peptide SEQ ID NO:14.  
 DE crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytosolic; antipsoriatic;  
 KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 1..15 /note= "optionally serially deleted"  
 FT Misc-difference 18..22 /note= "optionally serially deleted"  
 FT  
 XX WO2003012122-A2.  
 PN  
 XX 13-FEB-2003.  
 PD  
 XX 26-JUL-2002; 2002WO-US023763.  
 PF  
 XX 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0368854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX Sundaramoorthy M, Hudson B;  
 PI  
 XX

```

DR  WPI; 2003-332730/31.
XX
PT  New polypeptide, useful for treating an angiogenesis-mediated disease or
PT  condition consisting of glaucoma or blood-borne tumors or for inhibiting
PT  basal lamina membrane formation in cell or tissue development.
XX
XX  Claim 11; SEQ ID NO 14; 168pp; English.
PS
XX
CC  The present invention describes a crystallised NCI domain hexamer of type
CC  IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
CC  pharmaceutical composition comprising the polypeptide and a carrier; (3)
CC  inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
CC  disease or condition in a mammal; (5) inhibiting tumour metastasis or
CC  growth; (5) inhibiting endothelial cell interaction with the
CC  extracellular matrix in an animal tissue; (6) inhibiting basal lamina
CC  membrane formation in cell or tissue development; (7) a crystal of an NCI
CC  domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
CC  collagen assembly; and (9) an inhibitor of type IV collagen assembly. A
CC  crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
CC  antiproliferative, anti-anaemic, ophthalmological, antiarteriosclerotic and
CC  antiulcer activities, and can be used as an inhibitor of angiogenesis,
CC  tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
CC  cell proliferation, and basal lamina assembly. A (I) polypeptide can be
CC  used for treating an angiogenesis-mediated disease or condition
CC  consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
CC  psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
CC  or for inhibiting basal lamina membrane formation in cell or tissue
CC  development. The methods are useful for inhibiting angiogenesis in
CC  tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
CC  cell interaction with the extracellular matrix in an animal tissue, and
CC  identifying inhibitors of type IV collagen assembly. The present sequence
CC  represents a peptide which is used in the exemplification of the present
CC  invention.
XX
SQ  Sequence 22 AA;

Query Match      4.5%; Score 11; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  189 ECHGRGTCNY 199
DB  9 ECHGRGTCNY 19

RESULT 132
ADCI17415
ID  ADCI17415 standard; peptide; 22 AA.
XX
AC  ADCI17415;
XX
XX  18-DEC-2003 (first entry)
DT
DE  Type IV collagen NCI domain related peptide SEQ ID NO:16.
XX
KW  crystallised NCI domain hexamer of type IV collagen;
KW  angiogenesis inhibitor; angiogenesis-mediated disease;
KW  tumour metastasis inhibitor; tumour growth inhibitor;
KW  endothelial cell interaction inhibitor;
KW  basal lamina membrane formation inhibitor; cytostatic; antiproliferative;
KW  anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;
KW  endothelial cell adhesion inhibitor;
KW  endothelial cell proliferation inhibitor;
KW  ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;
KW  blood-borne tumour.
XX
OS  Synthetic.
OS  Homo sapiens.
XX
XX  Key Location/Qualifiers
PH
FT  Misc-difference 1..5
FT  /note= "optionally serially deleted"
FT  Misc-difference 18..22

```

```

FT  /note= "optionally serially deleted"
XX
XX  W02003012122-A2.
XX
XX  13-FEB-2003.
PD
XX
XX  26-JUL-2002; 2002WO-US023763.
XX
XX  27-JUL-2001; 2001US-0308523P.
XX  29-OCT-2001; 2001US-0351289P.
XX  22-MAR-2002; 2002US-0366854P.
XX  03-JUN-2002; 2002US-0385362P.
XX
XX  (UNIV ) UNIV KANSAS MEDICAL CENT.
XX  (SUND/) SUNDARAMOORTHY M.
XX  (HUDS/) HUDSON B.
XX
XX  Sundaramoorthy M, Hudson B;
XX  WPI; 2003-332730/31.
XX
XX  New polypeptide, useful for treating an angiogenesis-mediated disease or
XX  condition consisting of glaucoma or blood-borne tumors or for inhibiting
XX  basal lamina membrane formation in cell or tissue development.
XX
XX  Claim 11; SEQ ID NO 16; 168pp; English.
XX
XX  The present invention describes a crystallised NCI domain hexamer of type
XX  IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a
XX  pharmaceutical composition comprising the polypeptide and a carrier; (3)
XX  inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated
XX  disease or condition in a mammal; (5) inhibiting tumour metastasis or
XX  growth; (5) inhibiting endothelial cell interaction with the
XX  extracellular matrix in an animal tissue; (6) inhibiting basal lamina
XX  membrane formation in cell or tissue development; (7) a crystal of an NCI
XX  domain hexamer of type IV collagen; (8) identifying inhibitors of type IV
XX  collagen assembly; and (9) an inhibitor of type IV collagen (I) has cytostatic,
XX  crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,
XX  antiproliferative, anti-anaemic, ophthalmological, antiarteriosclerotic and
XX  antiulcer activities, and can be used as an inhibitor of angiogenesis,
XX  tumour growth, tumour metastasis, endothelial cell adhesion, endothelial
XX  cell proliferation, and basal lamina assembly. A (I) polypeptide can be
XX  used for treating an angiogenesis-mediated disease or condition
XX  consisting of glaucoma, sickle cell anaemia, ulcerative colitis,
XX  psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours
XX  or for inhibiting basal lamina membrane formation in cell or tissue
XX  development. The methods are useful for inhibiting angiogenesis in
XX  tissue, inhibiting tumour metastasis or growth, inhibiting endothelial
XX  cell interaction with the extracellular matrix in an animal tissue, and
XX  identifying inhibitors of type IV collagen assembly. The present sequence
XX  represents a peptide which is used in the exemplification of the present
XX  invention.
XX
SQ  Sequence 22 AA;

Query Match      4.5%; Score 11; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  189 ECHGRGTCNY 199
DB  9 ECHGRGTCNY 19

RESULT 133
ABB64070
ID  ABB64070 standard; protein; 1940 AA.
XX
XX  ABB64070;
AC
XX
XX  26-MAR-2002 (first entry)
DT
XX  Drosophila melanogaster polypeptide SEQ ID NO 19002.
DE

```

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
PD  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX N-PSDB; ABL08173.  
DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 19002; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1940 AA;  
SQ  
Query Match 4.5%; Score 11; DB 4; Length 1940;  
Best Local Similarity 100.0%; Pred. No. 0.092;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 174 SPGSCLEEFRA 184  
DB 1670 SPGSCLEEFRA 1680  
RESULT 134  
AAG93938  
ID AAG93938 standard; peptide; 10 AA.  
AC AAG93938;  
XX  
XX 18-SEP-2001 (first entry)  
DT  
DE Human complementary peptide; SEQ ID NO: 132.  
XX  
XX Human complementary peptide; ligand; drug discovery; drug design.  
XX Homo sapiens.  
OS  
XX WO200142277-A2.  
PN  
XX 14-JUN-2001.  
PD  
XX 13-DEC-2000; 2000WO-GB004776.  
PF  
XX 13-DEC-1999; 99GB-00029454.  
PR  
XX (PROT-) PROTEOM LTD.  
PA

XX Roberts GW, Heal JR;  
PI WPI; 2001-408419/43.  
DR  
XX A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
XX Example 4; Page 59; 646pp; English.  
PS  
XX The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
XX Sequence 10 AA;  
SQ  
Query Match 4.1%; Score 10; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.0082;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 158 IMFTSAGSEGE 167  
DB 1 IMFTSAGSEGE 10  
RESULT 135  
AAB97338  
ID AAB97338 standard; peptide; 10 AA.  
XX  
XX AAB97338;  
AC  
XX 13-AUG-2001 (first entry)  
DT  
DE Collagen IV alpha 3 domain epitope peptide #6.  
XX  
XX B cell; toxin; antigen specific; antibody mediated disease; virucide;  
KW immunosuppressive; anti-inflammatory; anti-allergic; antidiabetic;  
KW thyromimetic; antithyroid; vasotropic; cardiac; antitumor;  
KW neuroprotective; antirheumatic; antiarthritic; dermatological;  
KW ophthalmological; nephrotropic; allergic; autoimmune disorder;  
KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;  
KW cardiovascular disease; immunohaematologic disorder; neurologic disease;  
KW gastrointestinal disease; collagen vascular disease; renal diseases;  
KW pulmonary disease; infertility disorder; collagen IV;  
KW Goodpasture syndrome.  
XX  
XX Unidentified.  
OS  
XX WO200132853-A1.  
PN  
XX 10-MAY-2001.  
PD  
XX 12-OCT-2000; 2000WO-US028157.  
PF  
XX 29-OCT-1999; 99US-0162464P.  
PR  
XX (BIOM-) INST APPLIED BIOMEDICINE.  
XX  
XX Chaplin JW;  
PI  
XX WPI; 2001-316435/33.  
DR  
XX B cell clonal toxin useful for treating autoimmune disorders such as  
PT Grave's disease, myocardial infarction, Crohn's disease, multiple  
PT sclerosis, comprises a group that causes toxin to be internalized by B  
PT cell.  
XX  
XX Disclosure; Page 35; 46pp; English.  
PS

XX This invention relates to a B cell clonal toxin. The toxin is made from  
 CC two moieties, the first causes the toxin to be internalised by a B cell,  
 CC and the second is a biologically acceptable toxin. The invention includes  
 CC a method for inactivating/killing an antigen specific B cell. A target B  
 CC cell is contacted with an effective amount of a B cell clonal toxin. The  
 CC method is useful for selective immunosuppression in conditions  
 CC characterised by the presence of an unwanted or deleterious immune  
 CC response, e.g. in the treatment of antigen specific antibody mediated  
 CC disease conditions. Use of the B cell clonal toxin can result in  
 CC immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic  
 CC ; thyromimetic; antithyroid; vasotropic; cardiatic; antitumor;  
 CC neuroprotective; antirheumatic; antiarthritic; dermatological;  
 CC ophthalmological; and nephrotropic activity. The toxin is particularly  
 CC useful for treating a host suffering from an antigen specific antibody  
 CC mediated disease condition, where the antigen specific antibody is  
 CC produced by an antigen-reactive B cell population present in a host. The  
 CC toxin is useful for treating allergies, viral disease conditions, and  
 CC autoimmune disorders. Also treated are skin diseases; autoimmune  
 CC endocrinopathies; vasculitic syndromes; cardiovascular disease;  
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic  
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;  
 CC and infertility disorders. The present sequence represents a collagen IV  
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is  
 CC implicated in Goodpasture syndrome, a disorder which may be treated using  
 CC the toxin of the invention  
 XX Sequence 10 AA;

Query Match 4.1%; Score 10; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 FSLFLVQGNQ 54  
 |||||  
 Db 1 FSLFLVQGNQ 10

RESULT 136  
 ID AAB97336  
 AC AAB97336 standard; peptide; 10 AA.

XX AAB97336;

DT 13-AUG-2001 (first entry)

XX Collagen IV alpha 3 domain epitope peptide #4.

DE B cell; toxin; antigen specific; antibody mediated disease; virucide;  
 KW immunosuppressive; antiinflammatory; antiallergic; antidiabetic;  
 KW thyromimetic; antithyroid; vasotropic; cardiatic; antitumor;  
 KW neuroprotective; antirheumatic; antiarthritic; dermatological;  
 KW ophthalmological; nephrotropic; allergy; autoimmune disorder;  
 KW skin diseases; autoimmune endocrinopathy; vasculitic syndrome;  
 KW cardiovascular disease; immunohaematologic disorder; neurologic disease;  
 KW gastrointestinal disease; collagen vascular disease; renal diseases;  
 KW pulmonary disease; infertility disorder; collagen IV;  
 KW Goodpasture syndrome.

XX Unidentified.

XX WO200132853-A1.

XX 10-MAY-2001.

XX 12-OCT-2000; 2000WO-US028157.

XX 29-OCT-1999; 99US-0162464P.

XX (BIOM-) INST APPLIED BIOMEDICINE.

XX Chaplin JW;

XX

DR WPI; 2001-316435/33.

XX B cell clonal toxin useful for treating autoimmune disorders such as  
 PT Grave's disease, myocardial infarction, Crohn's disease, multiple  
 PT sclerosis, comprises a group that causes toxin to be internalized by B  
 PT cell.

XX Disclosure; Page 35; 46pp; English.

XX This invention relates to a B cell clonal toxin. The toxin is made from  
 CC two moieties, the first causes the toxin to be internalised by a B cell,  
 CC and the second is a biologically acceptable toxin. The invention includes  
 CC a method for inactivating/killing an antigen specific B cell. A target B  
 CC cell is contacted with an effective amount of a B cell clonal toxin. The  
 CC method is useful for selective immunosuppression in conditions  
 CC characterised by the presence of an unwanted or deleterious immune  
 CC response, e.g. in the treatment of antigen specific antibody mediated  
 CC disease conditions. Use of the B cell clonal toxin can result in  
 CC immunosuppressive; antiinflammatory; antiallergic; virucide; antidiabetic  
 CC ; thyromimetic; antithyroid; vasotropic; cardiatic; antitumor;  
 CC neuroprotective; antirheumatic; antiarthritic; dermatological;  
 CC ophthalmological; and nephrotropic activity. The toxin is particularly  
 CC useful for treating a host suffering from an antigen specific antibody  
 CC mediated disease condition, where the antigen specific antibody is  
 CC produced by an antigen-reactive B cell population present in a host. The  
 CC toxin is useful for treating allergies, viral disease conditions, and  
 CC autoimmune disorders. Also treated are skin diseases; autoimmune  
 CC endocrinopathies; vasculitic syndromes; cardiovascular disease;  
 CC immunohaematologic disorders; gastrointestinal diseases; neurologic  
 CC diseases; collagen vascular diseases; renal diseases; pulmonary diseases;  
 CC and infertility disorders. The present sequence represents a collagen IV  
 CC alpha 3 domain epitope peptide. An antibody response to this antigen is  
 CC implicated in Goodpasture syndrome, a disorder which may be treated using  
 CC the toxin of the invention  
 XX Sequence 10 AA;

Query Match 4.1%; Score 10; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LASLNPERMP 216  
 |||||

Db 1 LASLNPERMP 10

RESULT 137

ADA20257

ID ADA20257 standard; peptide; 10 AA.

XX ADA20257;

XX ADA20257;

DT 20-NOV-2003 (first entry)

DE Peptide Seq ID58 related to human type IV collagen and angiogenesis.  
 XX anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; human.

XX Homo sapiens.

XX WO2003059257-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US040938.

XX 21-DEC-2001; 2001US-00032221.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.



XX Kalluri R;  
 PI WPI; 2003-587256/55.  
 DR  
 XX  
 PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 XX  
 PS Claim 42; Page 174; 240pp; English.  
 XX  
 CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence peptide Seq IDS8, derived from the  
 CC sequence for human type IV collagen alpha 3 chain, which is preferably  
 CC including as part of a peptide of the invention.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 4.1%; Score 10; DB 6; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 88 ASRNDYSYWL 97  
 Db 1 ASRNDYSYWL 10  
 RESULT 138  
 ADC17678  
 ID ADC17678 standard; peptide; 10 AA.  
 XX  
 AC ADC17678;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Type IV collagen NC1 domain related peptide SEQ ID NO:283.  
 XX  
 KW crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
 KW blood-borne tumour.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FN WO2003012122-A2.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 26-JUL-2002; 2002WO-US023763.  
 XX  
 FR 27-JUL-2001; 2001US-0308523P.  
 FR 29-OCT-2001; 2001US-0351289P.

PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX  
 PI Sundaramoorthy M, Hudson B;  
 XX WPI; 2003-332730/31.  
 DR  
 XX  
 PT New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 XX  
 PS Claim 57; SEQ ID NO 283; 169pp; English.  
 XX  
 CC The present invention describes a crystallised NC1 domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NC1  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NC1 domain hexamer of type IV collagen (I) has cytostatic,  
 CC antipsoriatic, anti-anaemic, ophthalmological, antiarteriosclerotic and  
 CC antiulcer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 4.1%; Score 10; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 65 GSCIQRFETM 74  
 Db 1 GSCIQRFETM 10  
 RESULT 139  
 ADC17691  
 ID ADC17691 standard; peptide; 10 AA.  
 XX  
 AC ADC17691;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Type IV collagen NC1 domain related peptide SEQ ID NO:298.  
 XX  
 KW crystallised NC1 domain hexamer of type IV collagen;  
 KW angiogenesis inhibitor; angiogenesis-mediated disease;  
 KW tumour metastasis inhibitor; tumour growth inhibitor;  
 KW endothelial cell interaction inhibitor;  
 KW basal lamina membrane formation inhibitor; cytostatic; antipsoriatic;  
 KW anti-anaemic; ophthalmological; antiarteriosclerotic; antiulcer;  
 KW endothelial cell adhesion inhibitor;  
 KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
 KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;

KW blood-borne tumour.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO2003012122-A2.  
 XX  
 XX 13-FEB-2003.  
 XX  
 XX 26-JUL-2002; 2002WO-US023763.  
 XX  
 XX 27-JUL-2001; 2001US-0308523P.  
 PR 29-OCT-2001; 2001US-0351289P.  
 PR 22-MAR-2002; 2002US-0366854P.  
 PR 03-JUN-2002; 2002US-0385362P.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (SUND/) SUNDARAMOORTHY M.  
 PA (HUDS/) HUDSON B.  
 XX  
 PI Sundaramoorthy M, Hudson B;  
 XX  
 XX WPI; 2003-332730/31.  
 DR  
 XX  
 XX New polypeptide, useful for treating an angiogenesis-mediated disease or  
 PT condition consisting of glaucoma or blood-borne tumors or for inhibiting  
 PT basal lamina membrane formation in cell or tissue development.  
 PT  
 XX  
 PS Claim 57; SEQ ID NO 298; 168pp; English.  
 XX  
 CC The present invention describes a crystallised NCI domain hexamer of type  
 CC IV collagen (I). Also described: (1) a chimeric polypeptide; (2) a  
 CC pharmaceutical composition comprising the polypeptide and a carrier; (3)  
 CC inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
 CC disease or condition in a mammal; (5) inhibiting tumour metastasis or  
 CC growth; (5) inhibiting endothelial cell interaction with the  
 CC extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
 CC membrane formation in cell or tissue development; (7) a crystal of an NCI  
 CC domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
 CC collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
 CC crystallised NCI domain hexamer of type IV collagen (I) has cytostatic,  
 CC antiproliferative, antitumour, ophthalmological, antiarteriosclerotic and  
 CC anticancer activities, and can be used as an inhibitor of angiogenesis,  
 CC tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
 CC cell proliferation, and basal lamina assembly. A (I) polypeptide can be  
 CC used for treating an angiogenesis-mediated disease or condition  
 CC consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
 CC psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
 CC or for inhibiting basal lamina membrane formation in cell or tissue  
 CC development. The methods are useful for inhibiting angiogenesis in  
 CC tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
 CC cell interaction with the extracellular matrix in an animal tissue, and  
 CC identifying inhibitors of type IV collagen assembly. The present sequence  
 CC represents a peptide which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 4.1%; Score 10; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0052;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 78 FCVNDVCFNF 87  
 DB 1 FCVNDVCFNF 10  
 RESULT 140  
 ADA20239  
 ID ADA20239 standard; peptide; 27 AA.  
 XX  
 AC ADA20239;  
 XX

DT 20-NOV-2003 (first entry)  
 XX  
 DE T8-3 peptide related to human type IV collagen alpha and angiogenesis.  
 XX  
 KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
 KW metastasis; basement membrane organisation; type IV collagen network;  
 KW C-terminal globular non-collagenous domain; NCI; type IV collagen;  
 KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
 KW cytostatic; gene therapy; T8-3 peptide; tumstatin; human;  
 KW type IV collagen alpha 3 chain; mutant; mutein.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"  
 FT Misc-difference 12 /note= "Wild-type Cys substituted by Ser"  
 FT Misc-difference 18 /note= "Wild-type Cys substituted by Ser"  
 FT  
 XX  
 XX WO2003059257-A2.  
 XX  
 XX 24-JUL-2003.  
 PD  
 XX  
 XX 20-DEC-2002; 2002WO-US040938.  
 PF  
 XX  
 XX 21-DEC-2001; 2001US-00032221.  
 PR  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA  
 XX Kalluri R;  
 PI  
 XX WPI; 2003-587256/55.  
 DR  
 XX  
 PT New peptide, useful for preparing a composition for inhibiting tumor  
 PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
 PT  
 XX Claim 63; Page 45; 240pp; English.  
 PS  
 CC This invention relates to novel isolated proteins and their fragments  
 CC with anti-angiogenic properties. The invention also relates to the DNA  
 CC sequences which encode the novel proteins. A wide variety of diseases are  
 CC the result of undesirable angiogenesis. The formation of new capillaries  
 CC from pre-existing vessels is essential for tumour growth and metastasis.  
 CC Basement membrane organisation is dependent on the assembly of a type IV  
 CC collagen network which may occur through the C-terminal globular non-  
 CC collagenous (NCI) domain of type IV collagen. The alpha 1 and alpha 2  
 CC forms are ubiquitously exhibited in human basement membranes. In the  
 CC present invention, cell surface receptors (in particular integrins) which  
 CC specifically bind anti-angiogenic proteins and peptides (in particular  
 CC the alpha 1, alpha 2 and alpha 3 domains of the NCI domain of type IV  
 CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
 CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
 CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
 CC sequences of the invention may be useful in gene therapy. The present  
 CC sequence is the amino acid sequence of the T8-3 peptide of the invention,  
 CC derived from the amino acid sequence of tumstatin, which in turn was  
 CC derived from the amino acid sequence of human type IV collagen alpha 3  
 CC chain.  
 XX  
 SQ Sequence 27 AA;  
 Query Match 4.1%; Score 10; DB 6; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 69 QRFTMPPLF 78  
 DB 2 QRFTMPPLF 11

RESULT 141  
ADA20241  
ID ADA20241 standard; peptide; 27 AA.  
XX AC  
XX ADA20241;  
XX DT 20-NOV-2003 (first entry)  
XX DE P2 peptide related to human type IV collagen alpha and angiogenesis.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cytoskeletal; gene therapy; P2 peptide; tumstatin; human;  
XX KW type IV collagen alpha 3 chain; mutant; mutein.  
XX OS Synthetic.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 1 /note= "Wild-type Leu substituted by Lys"  
XX FT Misc-difference 12 /note= "Wild-type Cys substituted by Asp"  
XX FT Misc-difference 18 /note= "Wild-type Cys substituted by Asp"  
XX FT W02003059257-A2.  
XX PN 24-JUL-2003.  
XX PD 20-DEC-2002; 2002WO-US040938.  
XX PF 21-DEC-2001; 2001US-00032221.  
XX PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PA Kalluri R;  
XX PI WPI; 2003-587255/55.  
XX DR New peptide, useful for preparing a composition for inhibiting tumor  
XX PT growth, angiogenic activity or protein synthesis in a mammalian tissue.  
XX PS Claim 65; Page 45; 240pp; English.  
XX CC This invention relates to novel isolated proteins and their fragments  
XX CC with anti-angiogenic properties. The invention also relates to the DNA  
XX CC sequences which encode the novel proteins. A wide variety of diseases are  
XX CC the result of undesirable angiogenesis. The formation of new capillaries  
XX CC from pre-existing vessels is essential for tumour growth and metastasis.  
XX CC Basement membrane organisation is dependent on the assembly of a type IV  
XX CC collagen network which may occur through the C-terminal globular non-  
XX CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
XX CC forms are ubiquitously exhibited in human basement membranes. In the  
XX CC present invention, cell surface receptors (in particular integrins) which  
XX CC specifically bind anti-angiogenic proteins and peptides (in particular  
XX CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
XX CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
XX CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
XX CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
XX CC sequences of the invention may be useful in gene therapy. The present  
XX CC sequence is the amino acid sequence of the P2 peptide of the invention,  
XX CC derived from the amino acid sequence of tumstatin, which in turn was  
XX CC derived from the amino acid sequence of human type IV collagen alpha 3  
XX CC chain.  
XX SQ Sequence 27 AA;  
Query Match 4.1%; Score 10; DB 6; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 128 PAIAIAVHSQ 137  
DB 27 PAIAIAVHSQ 36

QY 69 QRETTMPFLF 78  
DB 2 QRETTMPFLF 11  
RESULT 142  
ABG70168  
ID ABG70168 standard; protein; 143 AA.  
XX AC ABG70168;  
XX DT 21-OCT-2002 (first entry)  
XX DE Human prey protein for Shigella ipaC #32.  
XX KW Prey protein; ospB; ospD1; ipaD; ipaC; ipaH9.8; ospG; ospC1; Shigella;  
XX KW shigellosis; bacillary dysentery; antibacterial; yeast two-hybrid system;  
XX KW protein-protein interaction; SID; selected interacting domain; human.  
XX OS Homo sapiens.  
XX OS W0200257303-A2.  
XX PN 25-JUL-2002.  
XX PD 11-JAN-2002; 2002WO-EP000777.  
XX PF 12-JAN-2001; 2001US-0261130P.  
XX PR (HYBR-) HYBRIGENICS.  
XX PA Legrain P;  
XX PI WPI; 2002-599706/64.  
XX DR N-PSDB; ABSS51561.  
XX PT New complex of protein-protein interactions between a bait Shigella  
XX PT flexneri polypeptide and a prey mammalian or human placenta polypeptide  
XX PT for treating or preventing bacillary dysentery in a mammal or human.  
XX PS Claim 7; Page 113; 162pp; English.  
XX CC The invention relates to a complex of protein-protein interactions  
XX CC between a Shigella flexneri polypeptide (e.g. ospB, ospD1, ipaD, ipaC,  
XX CC ipaH9.8, ospG and ospC1) and a mammalian polypeptide defined in the  
XX CC specification. The complexes are formed using the yeast two-hybrid  
XX CC system. Also included are (1) a recombinant host cell expressing the  
XX CC interactions between the Shigella flexneri polypeptide and a mammalian  
XX CC polypeptide defined in the specification; (2) selecting a modulating  
XX CC compound that inhibits or activates the protein-protein interactions; (3)  
XX CC a modulating compound obtained from the method of (2); (4) a SID  
XX CC (selected interacting domain) polypeptide or its fragment or variant  
XX CC comprising the human polypeptides appearing as ABG70042-ABG70242; (5) a  
XX CC SID polynucleotide or its fragment or variant comprising encoding the  
XX CC above polypeptides a vector comprising (5); (6) a recombinant host cell  
XX CC containing the vector; and (10) a protein chip comprising Shigella  
XX CC flexneri polypeptide and a mammalian polypeptide defined in the  
XX CC specification. A pharmaceutical composition comprising the compound,  
XX CC polypeptide or polynucleotide is useful for treating or preventing  
XX CC shigellosis (bacillary dysentery) in a human or mammal. The present  
XX CC sequence represents a human prey protein isolated by the yeast two-hybrid  
XX CC assay, forming a complex of the invention with a shigella protein  
XX SQ Sequence 143 AA;  
Query Match 4.1%; Score 10; DB 5; Length 143;  
Best Local Similarity 100.0%; Pred. No. 0.092;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 143

ADA20221  
ID ADA20221 standard; protein; 227 AA.

XX AC ADA20221;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human type IV collagen alpha 2 chain partial protein sequence.  
XX KW anti-angiogenic; undesirable angiogenesis; capillary; tumour growth;  
XX KW metastasis; basement membrane organisation; type IV collagen network;  
XX KW C-terminal globular non-collagenous domain; NC1; type IV collagen;  
XX KW cell surface receptor; integrin; angiogenic activity; protein synthesis;  
XX KW cystostatic; gene therapy; alpha 2 chain; canstatin; human.  
XX OS Homo sapiens.  
XX PN WO2003059257-A2.  
XX PD 24-JUL-2003.  
XX PF 20-DEC-2002; 2002WO-US040938.  
XX PR 21-DEC-2001; 2001US-00032221.  
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX PI Kalluri R;  
XX DR WPI; 2003-587256/55.  
XX DR N-PSDB; ADA20220.

PT New peptide, useful for preparing a composition for inhibiting tumor  
PT growth, angiogenic activity or protein synthesis in a mammalian tissue.

PS Claim 101; Fig 11; 240pp; English.

XX This invention relates to novel isolated proteins and their fragments  
CC with anti-angiogenic properties. The invention also relates to the DNA  
CC sequences which encode the novel proteins. A wide variety of diseases are  
CC the result of undesirable angiogenesis. The formation of new capillaries  
CC from pre-existing vessels is essential for tumour growth and metastasis.  
CC Basement membrane organisation is dependent on the assembly of a type IV  
CC collagen network which may occur through the C-terminal globular non-  
CC collagenous (NC1) domain of type IV collagen. The alpha 1 and alpha 2  
CC forms are ubiquitously exhibited in human basement membranes. In the  
CC present invention, cell surface receptors (in particular integrins) which  
CC specifically bind anti-angiogenic proteins and peptides (in particular  
CC the alpha 1, alpha 2 and alpha 3 domains of the NC1 domain of type IV  
CC collagen) are disclosed. The proteins of the invention may inhibit tumour  
CC growth, angiogenic activity in mammalian tissue or protein synthesis in  
CC endothelial cells and thus may exhibit cytostatic activity. The DNA  
CC sequences of the invention may be useful in gene therapy. The present  
CC sequence is the partial amino acid sequence of the alpha 2 chain of human  
CC type IV collagen. The "canstatin" peptide of the invention was derived  
CC from this protein.

XX Sequence 227 AA;

Query Match 4.1%; Score 10; DB 6; Length 227;

Best Local Similarity 100.0%; P-Value 0.14; Mismatches 0; Gaps 0;  
Matches 10; Conservative 0;

Oy 128 PAIAIAVHSQ 137

Db 111 PAIAIAVHSQ 120

## RESULT 144

ADC17696

ID ADC17696 standard; protein; 227 AA.

XX AC ADC17696;

XX DT 18-DEC-2003 (first entry)

XX DE Human type IV collagen alpha 2 chain protein SEQ ID NO:303.

XX KW crystallised NC1 domain hexamer of type IV collagen;  
XX KW angiogenesis inhibitor; angiogenesis-mediated disease;  
XX KW tumour metastasis inhibitor; tumour growth inhibitor;  
XX KW endothelial cell interaction inhibitor;  
XX KW basal lamina membrane formation inhibitor; cytostatic; antiproliferative;  
XX KW antiangiogenic; ophthalmological; antiarteriosclerotic; antiulcer;  
XX KW endothelial cell adhesion inhibitor;  
XX KW endothelial cell proliferation inhibitor; glaucoma; sickle cell anaemia;  
XX KW ulcerative colitis; psoriasis; atherosclerosis; rheumatoid arthritis;  
XX KW blood-borne tumour.

XX OS Homo sapiens.

XX PN WO2003012122-A2.

XX PD 13-FEB-2003.

XX PF 26-JUL-2002; 2002WO-US023763.

XX PR 27-JUL-2001; 2001US-0308523P.

XX PR 29-OCT-2001; 2001US-0351289P.

XX PR 22-MAR-2002; 2002US-0366854P.

XX PR 03-JUN-2002; 2002US-0385362P.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX (SUND/) SUNDARAMOORTHY M.

XX (HUDS/) HUDSON B.

XX Sundaramoorthy M, Hudson B;

XX WPI; 2003-332730/31.

XX PT New polypeptide, useful for treating an angiogenesis-mediated disease or  
XX condition consisting of glaucoma or blood-borne tumours or for inhibiting  
XX basal lamina membrane formation in cell or tissue development.

XX Disclosure; SEQ ID NO 303; 168pp; English.

XX The present invention describes a crystallised NC1 domain hexamer of type  
XX IV collagen (1). Also described: (1) a chimeric polypeptide; (2) a  
XX pharmaceutical composition comprising the polypeptide and a carrier; (3)  
XX inhibiting angiogenesis in tissue; (4) treating an angiogenesis-mediated  
XX disease or condition in a mammal; (5) inhibiting tumour metastasis or  
XX growth; (5) inhibiting endothelial cell interaction with the  
XX extracellular matrix in an animal tissue; (6) inhibiting basal lamina  
XX membrane formation in cell or tissue development; (7) a crystal of an NC1  
XX domain hexamer of type IV collagen; (8) identifying inhibitors of type IV  
XX collagen assembly; and (9) an inhibitor of type IV collagen assembly. A  
XX crystallised NC1 domain hexamer of type IV collagen (1) has cytostatic,  
XX antiproliferative, antiangiogenic, ophthalmological, antiarteriosclerotic and  
XX antiulcer activities, and can be used as an inhibitor of angiogenesis,  
XX tumour growth, tumour metastasis, endothelial cell adhesion, endothelial  
XX cell proliferation, and basal lamina assembly. A (1) polypeptide can be  
XX used for treating an angiogenesis-mediated disease or condition  
XX consisting of glaucoma, sickle cell anaemia, ulcerative colitis,  
XX psoriasis, atherosclerosis, rheumatoid arthritis or blood-borne tumours  
XX or for inhibiting basal lamina membrane formation in cell or tissue  
XX development. The methods are useful for inhibiting angiogenesis in  
XX tissue, inhibiting tumour metastasis or growth, inhibiting endothelial  
XX cell interaction with the extracellular matrix in an animal tissue, and  
XX identifying inhibitors of type IV collagen assembly. The present sequence  
XX represents an amino acid sequence which is used in the exemplification of  
XX the present invention.

XX Sequence 227 AA;

Query Match 4.1%; Score 10; DB 7; Length 227;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
DB 111 PAIAIAVHSQ 120

RESULT 145  
AAU75588  
ID AAU75588 standard; protein; 241 AA.  
AC AAU75588;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human type IV collagen alpha 2 chain.  
XX  
KW Human; type IV collagen alpha 2 chain; cytostatic; antiangiogenic;  
KW non-Goodpasture fragment; alpha3(IV)NC1 domain; alphavbeta3 integrin;  
KW endothelial cell proliferation; apoptosis; Arresten; Canstatin;  
KW Tumstatin; angiogenesis; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 166 /note= "Encoded by gtc"  
FT  
FN WO200151523-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 08-JAN-2001; 2001WO-US000565.  
XX  
PR 07-JAN-2000; 2000US-00479118.  
PR 04-APR-2000; 2000US-00543371.  
PR 21-JUL-2000; 2000US-00625191.  
XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
PI Kalluri R;  
XX  
DR WPI; 2002-188037/24.  
DR N-PSDB; ABK15362.  
XX  
PT A non-Goodpasture fragment of alpha3(IV)NC1 domain used in detecting and  
XX treating disorders involving angiogenesis.  
PS Example 14; Fig 11B; 205pp; English.

The invention relates to a non-Goodpasture fragment of alpha3(IV)NC1 domain, having one or more of the characteristics selected from: (a) the ability to bind alphavbeta3 integrin; (b) the ability to inhibit proliferation of endothelial cells; and (c) the ability to cause apoptosis of endothelial cells. Also described are the following: (1) use of Arresten, Canstatin or Tumstatin, or a fragment, mutant, homologue, analogue or allelic variant in the preparation of a medicament for treating a disorder involving: (a) inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or (b) by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; (2) use of an antibody or peptide that specifically binds the alpha1, alpha2, alpha3, alpha5, alpha6, alpha7, beta1 or beta3 subunit of integrin in the preparation of a medicament for inhibiting angiogenesis or cell proliferation; (3) use of an inhibitor, such as an antibody, antibody fragment or peptide of receptor-mediated angiogenesis in the preparation of a medicament for treating a proliferative disease in a vertebrate, where the disease is characterised by angiogenesis that is mediated by

receptors to Arresten, Canstatin or Tumstatin and where the receptors inhibited are Arresten, Canstatin or Tumstatin receptors; (4) use of one or more soluble receptors that bind Arresten, Canstatin or Tumstatin in the presence of a medicament for promoting angiogenesis in a tissue; and (5) use of integrins in the preparation of a medicament for promoting or inducing angiogenesis or cell proliferation in a tissue. The fragments Arresten, Canstatin or Tumstatin and their mutants, homologues, analogues or allelic variants are useful in the preparation of a medicament for treating a disorder involving inhibiting angiogenesis in a tissue, where the angiogenesis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits; or by promoting or inducing endothelial cell apoptosis in a tissue, where the endothelial cell apoptosis is mediated by one or more endothelial cell integrins or one or more endothelial cell integrin subunits. The medicament is useful in inhibiting tumour growth and for the regression of an established tumour. The present sequence represents the amino acid sequence of human type IV collagen alpha 2 chain

Sequence 241 AA;  
Query Match 4.1%; Score 10; DB 5; Length 241;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
DB 126 PAIAIAVHSQ 135

RESULT 146  
AAU67946  
ID AAU67946 standard; protein; 242 AA.  
XX  
AC AAU67946;  
XX  
DT 03-APR-2000 (first entry)  
XX  
DE Human type IV collagen alpha 2 chain protein sequence SEQ ID NO:6.  
XX  
KW Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;  
KW benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;  
KW ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;  
KW myocardial angiogenesis; plaque neovascularisation; angiodioma;  
KW atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;  
KW Contraception; Obesity.  
XX  
OS Homo sapiens.  
XX  
PN WO9965940-A1.  
XX  
PD 23-DEC-1999.  
XX  
PF 17-JUN-1999; 99WO-US013737.  
XX  
PR 17-JUN-1998; 98US-0089689P.  
PR 25-MAR-1999; 99US-0126175P.  
XX  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
PI Kalluri R;  
XX  
DR WPI; 2000-037708/08.  
DR N-PSDB; AA257162.  
XX  
PT Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or 3  
XX Chain of Type IV collagen used in, e.g. treatment of benign tumors and  
XX rheumatoid arthritis.  
PS Example 11; Fig 10B; 117pp; English.  
XX  
CC The present sequence represents the human type IV collagen alpha 2 chain.  
CC The present invention describes an isolated protein chosen from the NC1  
CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or a

CC fragment, analogue, derivative or mutant, which has anti-angiogenic  
 CC properties. The anti-angiogenic proteins, multimers and chimeras are  
 CC useful for inhibiting angiogenic activity in mammalian tissue, especially  
 CC for treating diseases chosen from angiogenesis-dependent cancers, benign  
 CC tumours, rheumatoid arthritis, diabetic retinopathy, psoriasis, ocular  
 CC angiogenesis diseases, Osler-Weber Syndrome, myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter pylori  
 CC ulcers, dialysis graft vascular access stenosis, contraception and  
 CC obesity. The compositions can be used to inhibit a disease characterised  
 CC by angiogenic activity, in conjunction with radiation therapy,  
 CC chemotherapy or immunotherapy  
 XX  
 XX Sequence 242 AA;

Query Match 4.1%; Score 10; DB 3; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137

DB 126 PAIAIAVHSQ 135

RESULT 147

AAV31992

ID AAV31992 standard; protein; 258 AA.

AC AAV31992;

XX

XX 05-JAN-2000 (first entry)

DE Type IV collagen NC1 domain alpha-2 monomer.

XX Type IV collagen; NC1 domain; non-collagenous domain; human;  
 KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
 KW rheumatoid arthritis; retinal neovascularization;  
 KW chorioidal neovascularization; macular degeneration;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW epidemic keratoconjunctivitis; vitamin A deficiency;  
 KW contact lens overwear; atopic keratitis; superior limbic keratitis;  
 KW pterygium keratitis sicca; sogrens; acne rosacea; phlyctenulosis;  
 KW syphilis; Mycobacteria infection; lipid degeneration; chemical burn;  
 KW ulcer; herpes simplex infection; Herpes zoster infection;  
 KW protozoan infection; Kaposi's sarcoma; Mooren ulcer;  
 KW Terrien's marginal degeneration; marginal keratolysis; trauma;  
 KW systemic lupus; polyarteritis; Wegener's sarcoidosis; scleritis;  
 KW Steven's Johnson disease; radial keratotomy; sickle cell anemia;  
 KW sarcoïdosis; pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease; myopia;  
 KW optic pit; Stargart's disease; pars planitis; chronic retinal detachment;  
 KW hyperviscosity syndrome; toxoplasmosis; post-laser complication;  
 KW fibrovascular tissue proliferation; haemangioma; Osler-Weber-Rendu; AIDS;  
 KW ocular neovascular disease; osteoarthritis; chronic inflammation;  
 KW Crohn's disease; ulcerative colitis; psoriasis; atherosclerosis;  
 KW pemphigoid.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..17

FT Peptide /note= "BM40 signal peptide"

FT Protein 18..258

FT Peptide /note= "mature protein"

FT Peptide 18..25

FT Protein /note= "affinity tag"

FT Protein 26..258

FT Protein /note= "NC1 alpha-2 monomer"

XX

PN WO9949885-A2.  
 XX  
 PD 07-OCT-1999.  
 XX  
 PF 26-MAR-1999; 99WO-US006445.  
 XX  
 PR 27-MAR-1998; 98US-0079783P.  
 XX  
 PR 29-OCT-1998; 98US-0106170P.  
 XX  
 PA (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Sarra MP;

XX WPI; 1999-601297/51.

XX DR N-PSDB; AA220090.

XX

PT Inhibition of angiogenesis with non-collagenous alpha chain monomer  
 PT useful for treating e.g. tumor growth or metastasis, neovascularisation,  
 PT etc.

XX Disclosure; Fig 17b; 56pp; English.

XX This sequence represents a recombinant type IV collagen non-collagenous  
 CC (NC1) domain alpha-2 polypeptide composed of a BM40 signal sequence  
 CC (which is cleaved from the mature protein) to facilitate protein  
 CC secretion, and a mature protein comprising an affinity tag (facilitates  
 CC purification and identification of the material) and the alpha-1 chain  
 CC monomer. The invention provides methods and kits for inhibiting  
 CC angiogenesis, tumour growth and metastasis, and endothelial cell  
 CC interacting with the extracellular matrix, each method comprising  
 CC contacting the tumour or animal tissue with 1 or more isolated type IV  
 CC collagen NC1 alpha chain monomer(s) selected from the group consisting of  
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-  
 CC 96). The monomers can be produced via recombinant protein expression. The  
 CC polynucleotides and polypeptides are used to treat an angiogenesis-  
 CC mediated disorder or condition, especially selected from solid and blood-  
 CC borne tumours, diabetic retinopathy, rheumatoid arthritis, retinal  
 CC neovascularization, chorioidal neovascularization, macular degeneration,  
 CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia, epidemic  
 CC keratoconjunctivitis, vitamin A deficiency, contact lens overwear, atopic  
 CC keratitis, superior limbic keratitis, pterygium keratitis sicca, sogrens,  
 CC acne rosacea, phlyctenulosis, syphilis, mycobacteria infections, lipid  
 CC degeneration, chemical burns, bacterial ulcers, fungal ulcers, herpes  
 CC simplex infections, herpes zoster infections, protozoan infections,  
 CC Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal  
 CC keratolysis, trauma, systemic lupus, polyarteritis, Wegener's  
 CC sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy,  
 CC sickle cell anemia, sarcoïd, pseudoxanthoma elasticum, Pagets disease,  
 CC vein occlusion, artery occlusion, carotid obstructive disease, chronic  
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets  
 CC disease, myopia, optic pits, Stargarts disease, pars planitis, chronic  
 CC retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser  
 CC complications, abnormal proliferation of fibrovascular tissue,  
 CC haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular disease,  
 CC osteoarthritis, chronic inflammation, Crohn's disease, ulcerative  
 CC colitis, psoriasis, atherosclerosis, and pemphigoid (all claimed)

XX Sequence 258 AA;

SQ

Query Match 4.1%; Score 10; DB 2; Length 258;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137

DB 142 PAIAIAVHSQ 151

RESULT 148

RAY97554

ID AAY97554 standard; protein; 258 AA.

XX

AC AAY97554;  
 XX 12-FEB-2001 (first entry)  
 XX Human alpha2(IV)NC1 protein sequence.  
 XX  
 XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;  
 XX tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;  
 XX retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukemia;  
 XX diabetic retinopathy; rheumatoid arthritis; neovascularisation;  
 XX muscular degeneration; corneal graft rejection; vitamin A deficiency;  
 XX atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;  
 XX Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;  
 XX chronic inflammation; psoriasis; therapy; alpha2(IV)NC1.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200059532-A1.  
 XX  
 XX 12-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US008678.  
 XX  
 XX 01-APR-1999; 99US-0127391P.  
 XX (BIOS-) BIOSTRATUM INC.  
 XX  
 XX Brooks P, Hudson B;  
 XX  
 XX WPI; 2000-664962/64.  
 XX N-PSDB; AAA90992.  
 XX  
 XX Use of antagonists of specific integrin receptors for inhibiting  
 XX angiogenesis, tumor growth or metastases, or endothelial cell  
 XX interactions with the extracellular matrix.  
 XX  
 XX Disclosure; Fig 17b; 78pp; English.  
 XX  
 XX This sequence is a human type IV collagen alpha chain monomer, designated  
 XX alpha2(IV)NC1. The invention relates to a method for inhibiting  
 XX angiogenesis, tumor growth or metastases, or endothelial cell  
 XX interactions with the extracellular matrix, comprising contacting the  
 XX cells or tissue with a polypeptide composition containing antagonists of  
 XX specific integrin receptors. The methods and the antagonists are useful  
 XX for inhibiting angiogenesis, tumor growth or metastases, or endothelial  
 XX cell interaction with the extracellular matrix. The antagonists are also  
 XX useful for treating diseases and conditions with accompanying undesired  
 XX angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,  
 XX carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,  
 XX neuroblastoma, osteosarcoma or leukemia). These are also applicable to  
 XX treating non-tumorigenic diseases and conditions with accompanying  
 XX undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,  
 XX retinal neovascularisation, choroidal neovascularisation, muscular  
 XX degeneration, corneal graft rejection, vitamin A deficiency, atopic  
 XX keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,  
 XX sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser  
 XX complications, chronic inflammation or psoriasis  
 XX  
 SQ Sequence 258 AA;  
 Query Match 4.1%; Score 10; DB 3; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 0.16;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 PAIAIAVHSQ 137  
 DB 142 PAIAIAVHSQ 151  
 RESULT 149  
 AAB58180  
 ID AAB58180 standard; protein; 430 AA.  
 XX

AC AAB58180;  
 XX 14-MAR-2001 (first entry)  
 XX Lung cancer associated polypeptide sequence SEQ ID 518.  
 XX  
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
 XX cardioactive; immunomodulatory; muscular active; vulnerary;  
 XX gastrointestinal; nephrotoxic; anti-infective; gynecological;  
 XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 XX proliferative disorder; wound healing; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200055180-A2.  
 XX  
 XX 21-SEP-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US005918.  
 XX  
 XX 12-MAR-1999; 99US-0124270P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX  
 XX Ruben SM;  
 XX  
 XX WPI; 2000-587514/55.  
 XX N-PSDB; AAF18056.  
 XX  
 XX Lung cancer associated gene sequences, referred to as lung cancer  
 XX antigens, useful for treatment, prevention, and diagnosis of disorders  
 XX such as lung cancer.  
 XX  
 XX Claim 11; Page 1008-1010; 1425pp; English.  
 XX  
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 XX associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 XX associated proteins and polynucleotide sequences, their agonists, and  
 XX antagonists may have neuroprotective, cytostatic, cardioactive, and  
 XX immunomodulatory, muscular active general; vulnerary; gastrointestinal  
 XX general; nephrotoxic; anti-infective; gynecological; or antibacterial  
 XX activity. The invention also includes antibodies specific for the protein  
 XX or polynucleotide sequences. The lung cancer associated polynucleotide  
 XX sequences may be used for detection of lung cancer, chromosome  
 XX identification, as chromosome markers, and for numerous other diagnostic  
 XX or research purposes. The proteins may be used to treat disorders such as  
 XX neural, immune, muscular, reproductive, gastrointestinal, pulmonary,  
 XX cardiovascular, renal, and proliferative disorders. The proteins may also  
 XX be used in the treatment of wounds and infectious diseases.  
 XX Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are  
 XX used in the course of the invention for the identification and  
 XX characterisation of the polynucleotide and protein sequences  
 XX  
 SQ Sequence 430 AA;  
 Query Match 4.1%; Score 10; DB 3; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 0.25;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 PAIAIAVHSQ 137  
 DB 314 PAIAIAVHSQ 323  
 RESULT 150  
 ADD89023  
 ID ADD89023 standard; protein; 459 AA.  
 XX  
 XX ADD89023;  
 XX  
 XX 29-JAN-2004 (first entry)  
 XX



```

DE TAT263.
XX tumour-associated antigenic target polypeptide; Cytostatic; tumour;
KW cancer.
XX
XX OS Homo sapiens.
XX PN WO2003057160-A2.
XX PD 17-JUL-2003.
XX
XX 30-DEC-2002; 2002WO-US041798.
XX
XX 02-JAN-2002; 2002US-0345444P.
XX PR 25-JAN-2002; 2002US-0351885P.
XX PR 25-FEB-2002; 2002US-0360066P.
XX PR 05-MAR-2002; 2002US-0362004P.
XX PR 20-MAR-2002; 2002US-0366869P.
XX PR 21-MAR-2002; 2002US-0366284P.
XX PR 28-MAR-2002; 2002US-0368679P.
XX PR 19-AUG-2002; 2002US-0404809P.
XX PR 21-AUG-2002; 2002US-0405645P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Frantz G, Hillan KJ, Phillips H, Polakis P, Smith V, Spencer SD;
XX Williams PM, Wu TD, Zhang Z;
XX
XX WPI; 2003-569537/53.
XX DR N-PSDB; ADD89098.
XX
XX PT New antibodies against tumor-associated antigenic target polypeptide,
XX useful for treating or diagnosing tumors or cancers in mammals, e.g.
XX prostate cancer, lung cancer, prostate adenocarcinomas or renal cell
XX carcinomas.
XX
XX Claim 1; SEQ ID NO 27; 252pp; English.
XX
XX PS The present invention relates to antibodies against tumour-associated
XX antigenic target polypeptide. The antibody is useful for treating or
XX diagnosing tumors or cancers in mammals, e.g. prostate cancer, lung
XX cancer, breast cancer, colon cancer, ovarian cancer, prostate
XX adenocarcinomas, renal cell carcinomas, or pleural mesothelioma. The
XX present sequence represents a TAT polypeptide.
XX
XX SQ Sequence 459 AA;
XX
XX Query Match 4.1%; Score 10; DB 7; Length 459;
XX Best Local Similarity 100.0%; Pred. NO. 0.26;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 128 PAIAIAVHSQ 137
XX Db 343 PAIAIAVHSQ 352
XX
XX Search completed: April 5, 2004, 07:37:22
XX Job time : 64 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:38:59 ; Search time 45 Seconds  
(without alignments)  
1423.863 Million cell updates/sec

Title: US-10-032-221B-10

Perfect score: 244

Sequence: 1 GLKGRDGSPTWTRGF.....KAGELEKISRCQVCMKKRH 244

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1071436 seqs, 262597696 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1071436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	100.0	244	14	US-10-032-221B-10
2	191	78.3	191	14	US-10-032-221B-22
3	159	65.2	211	14	US-10-032-221B-46
4	159	65.2	211	14	US-10-032-221B-46
5	155	63.5	232	14	US-10-032-221B-304
6	132	54.1	132	14	US-10-032-221B-23
7	124	50.8	124	14	US-10-032-221B-20
8	120	49.2	120	14	US-10-032-221B-21
9	112	45.9	112	14	US-10-032-221B-24
10	88	36.1	88	14	US-10-032-221B-33
11	80	32.8	88	14	US-10-032-221B-34
12	79	32.4	79	14	US-10-032-221B-26
13	64	28.2	64	14	US-10-032-221B-25
14	61	25.0	68	14	US-10-032-221B-50
15	61	25.0	68	14	US-10-032-221B-50

25.0	72	14	US-10-032-221B-48	Sequence 48, Appl
25.0	72	14	US-10-032-221B-52	Sequence 52, Appl
25.0	72	14	US-10-032-221B-52	Sequence 52, Appl
25.0	72	14	US-10-032-221B-52	Sequence 52, Appl
15.2	72	14	US-10-032-221B-61	Sequence 61, Appl
15.2	72	14	US-10-032-221B-39	Sequence 39, Appl
10.7	27	14	US-10-032-221B-37	Sequence 37, Appl
10.2	25	14	US-10-032-221B-15	Sequence 15, Appl
9.0	22	14	US-10-032-221B-15	Sequence 15, Appl
8.6	21	14	US-10-032-221B-26	Sequence 26, Appl
8.6	21	14	US-10-032-221B-26	Sequence 26, Appl
8.6	21	14	US-10-032-221B-26	Sequence 26, Appl
8.2	20	14	US-10-032-221B-28	Sequence 28, Appl
8.2	20	14	US-10-032-221B-29	Sequence 29, Appl
8.2	20	14	US-10-032-221B-30	Sequence 30, Appl
8.2	20	14	US-10-032-221B-27	Sequence 27, Appl
7.8	19	14	US-10-032-221B-31	Sequence 31, Appl
7.8	19	14	US-10-032-221B-32	Sequence 32, Appl
7.4	18	14	US-10-032-221B-254	Sequence 254, Appl
7.4	18	14	US-10-032-221B-260	Sequence 260, Appl
7.4	18	14	US-10-032-221B-277	Sequence 277, Appl
7.0	16	14	US-09-864-741-48095	Sequence 48095, A
7.0	16	14	US-10-032-221B-302	Sequence 302, Appl
7.0	16	14	US-10-032-221B-306	Sequence 306, Appl
7.0	16	14	US-10-032-221B-42	Sequence 42, Appl
7.0	16	14	US-09-925-297-496	Sequence 496, Appl
7.0	16	14	US-09-925-302-507	Sequence 507, Appl
7.0	16	14	US-10-032-221B-8	Sequence 8, Appl
6.6	15	14	US-10-032-221B-74	Sequence 74, Appl
6.1	15	14	US-10-032-221B-53	Sequence 53, Appl
6.1	15	14	US-10-032-221B-94	Sequence 94, Appl
6.1	15	14	US-10-032-221B-137	Sequence 137, Appl
6.1	15	14	US-10-032-221B-191	Sequence 191, Appl
6.1	15	14	US-10-032-221B-212	Sequence 212, Appl
6.1	15	14	US-10-032-221B-231	Sequence 231, Appl
5.7	14	14	US-10-032-221B-28	Sequence 28, Appl
5.7	14	14	US-10-032-221B-112	Sequence 112, Appl
5.7	14	14	US-10-032-221B-165	Sequence 165, Appl
5.7	14	14	US-10-032-221B-272	Sequence 272, Appl
5.7	14	14	US-10-032-221B-210	Sequence 210, Appl
5.7	21	14	US-10-032-221B-27	Sequence 27, Appl
5.7	21	14	US-10-032-221B-38	Sequence 38, Appl
5.7	25	14	US-10-032-221B-10	Sequence 10, Appl
4.9	12	14	US-10-032-221B-150	Sequence 150, Appl
4.5	11	14	US-10-032-221B-176	Sequence 176, Appl
4.5	11	14	US-10-032-221B-239	Sequence 239, Appl
4.5	11	14	US-10-032-221B-271	Sequence 271, Appl
4.5	19	14	US-10-032-221B-41	Sequence 41, Appl
4.5	22	14	US-10-032-221B-14	Sequence 14, Appl
4.5	22	14	US-10-032-221B-16	Sequence 16, Appl
4.5	1744	15	US-10-032-221B-5832	Sequence 5832, Ap
4.1	10	14	US-09-572-404B-132	Sequence 132, Ap
4.1	10	14	US-10-032-221B-283	Sequence 283, Appl
4.1	10	14	US-10-032-221B-298	Sequence 298, Appl
4.1	10	14	US-10-032-221B-58	Sequence 58, Appl
4.1	27	14	US-10-032-221B-40	Sequence 40, Appl
4.1	27	14	US-10-032-221B-42	Sequence 42, Appl
4.1	143	14	US-10-032-221B-342	Sequence 342, Appl
4.1	227	14	US-10-032-221B-303	Sequence 303, Appl
4.1	227	14	US-10-032-221B-6	Sequence 6, Appl
4.1	430	9	US-09-925-302-518	Sequence 518, Appl
4.1	459	15	US-10-331-496A-27	Sequence 27, Appl
4.1	459	15	US-10-372-683-30	Sequence 30, Appl
4.1	1712	10	US-09-961-403-9	Sequence 9, Appl
3.7	12	14	US-10-032-221B-99	Sequence 99, Appl
3.7	16	14	US-10-032-221B-76	Sequence 76, Appl
3.7	18	14	US-10-032-221B-259	Sequence 259, Appl
3.7	18	14	US-10-032-221B-261	Sequence 261, Appl
3.7	22	14	US-10-032-221B-265	Sequence 265, Appl

89 9 3.7 22 14 US-10-206-699-267 Sequence 267, App  
90 8 3.3 8 14 US-10-032-221B-50 Sequence 50, Appl  
91 8 3.3 8 14 US-10-032-221B-56 Sequence 56, Appl  
92 8 3.3 14 14 US-10-032-221B-44 Sequence 44, Appl  
93 8 3.3 14 14 US-10-270-877-44 Sequence 44, Appl  
94 8 3.3 18 14 US-10-206-699-253 Sequence 253, Appl  
95 8 3.3 19 14 US-10-270-877-43 Sequence 43, Appl  
96 8 3.3 19 14 US-10-270-877-43 Sequence 43, Appl  
97 8 3.3 20 14 US-10-206-699-288 Sequence 288, Appl  
98 8 3.3 228 14 US-10-206-699-307 Sequence 307, Appl  
99 8 3.3 937 15 US-10-369-493-3337 Sequence 3337, Ap  
100 7 2.9 7 14 US-10-032-221B-49 Sequence 49, Appl  
101 7 2.9 7 14 US-10-032-221B-55 Sequence 55, Appl  
102 7 2.9 8 14 US-10-032-221B-51 Sequence 51, Appl  
103 7 2.9 14 14 US-10-206-699-26 Sequence 26, Appl  
104 7 2.9 15 14 US-10-206-699-92 Sequence 92, Appl  
105 7 2.9 15 14 US-10-206-699-230 Sequence 230, Appl  
106 7 2.9 15 14 US-10-206-699-232 Sequence 232, Appl  
107 7 2.9 70 9 US-09-864-761-37448 Sequence 37448, A  
108 7 2.9 70 9 US-09-864-761-47938 Sequence 47938, A  
109 7 2.9 78 12 US-10-424-599-243322 Sequence 243322,  
110 7 2.9 84 12 US-10-424-599-182069 Sequence 182069,  
111 7 2.9 87 9 US-09-764-877-1521 Sequence 1521, Ap  
112 7 2.9 87 15 US-10-242-515-1521 Sequence 1521, Ap  
113 7 2.9 111 12 US-10-425-114-54455 Sequence 54455, A  
114 7 2.9 117 11 US-09-461-580A-18 Sequence 18, Appl  
115 7 2.9 142 9 US-09-864-761-38021 Sequence 38021, A  
116 7 2.9 150 15 US-10-094-749-3238 Sequence 3238, Ap  
117 7 2.9 174 12 US-10-425-114-47774 Sequence 47774, A  
118 7 2.9 198 9 US-09-727-855B-7 Sequence 7, Appl  
119 7 2.9 218 12 US-10-425-114-59362 Sequence 59362, A  
120 7 2.9 221 12 US-10-267-682-96 Sequence 96, Appl  
121 7 2.9 221 12 US-10-267-748-96 Sequence 96, Appl  
122 7 2.9 231 14 US-10-206-699-305 Sequence 305, Appl  
123 7 2.9 233 14 US-10-230-331-28 Sequence 28, Appl  
124 7 2.9 247 12 US-10-282-122A-72960 Sequence 72960, A  
125 7 2.9 251 12 US-10-282-122A-76032 Sequence 76032, A  
126 7 2.9 251 12 US-10-296-115-808 Sequence 808, App  
127 7 2.9 300 12 US-10-424-599-235900 Sequence 235900,  
128 7 2.9 304 12 US-09-964-956-71 Sequence 71, Appl  
129 7 2.9 311 12 US-10-425-114-59795 Sequence 59795, A  
130 7 2.9 314 12 US-10-425-114-71839 Sequence 71839, A  
131 7 2.9 317 12 US-10-425-114-50147 Sequence 50147, A  
132 7 2.9 328 12 US-10-424-599-201843 Sequence 201843,  
133 7 2.9 347 9 US-09-918-568-58 Sequence 58, Appl  
134 7 2.9 352 9 US-09-860-351-2 Sequence 2, Appl  
135 7 2.9 352 15 US-10-353-690-44 Sequence 44, Appl  
136 7 2.9 372 9 US-09-973-963-4 Sequence 4, Appl  
137 7 2.9 372 9 US-09-973-064-4 Sequence 4, Appl  
138 7 2.9 372 9 US-09-973-077-4 Sequence 4, Appl  
139 7 2.9 372 9 US-09-973-063-4 Sequence 4, Appl  
140 7 2.9 372 9 US-09-973-964-4 Sequence 4, Appl  
141 7 2.9 372 9 US-09-975-072-4 Sequence 4, Appl  
142 7 2.9 372 9 US-09-972-038-4 Sequence 4, Appl  
143 7 2.9 372 9 US-09-972-757-4 Sequence 4, Appl  
144 7 2.9 372 9 US-09-973-965-4 Sequence 4, Appl  
145 7 2.9 372 9 US-09-973-941-4 Sequence 4, Appl  
146 7 2.9 372 10 US-09-986-992-2 Sequence 2, Appl  
147 7 2.9 372 10 US-09-971-782-4 Sequence 4, Appl  
148 7 2.9 372 15 US-10-094-749-1699 Sequence 1699, Ap  
149 7 2.9 372 16 US-10-311-764-1 Sequence 1, Appl  
150 7 2.9 373 9 US-09-925-300-1655 Sequence 1655, Ap  
151 7 2.9 383 9 US-09-147-346-2 Sequence 2, Appl  
152 7 2.9 386 12 US-10-425-114-59796 Sequence 59796, A  
153 7 2.9 404 12 US-10-424-599-150766 Sequence 150766,  
154 7 2.9 421 14 US-10-103-313-450 Sequence 450, App  
155 7 2.9 423 12 US-10-282-122A-73255 Sequence 73255, A  
156 7 2.9 424 12 US-10-425-114-40411 Sequence 40411, A  
157 7 2.9 469 12 US-10-425-114-64052 Sequence 178, App  
158 7 2.9 478 11 US-09-906-179A-178 Sequence 230, App  
159 7 2.9 478 11 US-09-906-179A-230 Sequence 142, App  
160 7 2.9 478 12 US-10-671-403-142 Sequence 142, App  
161 7 2.9 478 12 US-10-671-419-142 Sequence 142, App

162 7 2.9 478 12 US-10-670-844-142 Sequence 142, App  
163 7 2.9 478 12 US-10-671-134-142 Sequence 142, App  
164 7 2.9 478 12 US-10-673-098-142 Sequence 142, App  
165 7 2.9 488 12 US-10-424-599-198256 Sequence 198256,  
166 7 2.9 490 12 US-10-425-114-47211 Sequence 47211, A  
167 7 2.9 503 14 US-10-096-840D-2 Sequence 2, Appl  
168 7 2.9 503 14 US-10-065-133A-8 Sequence 8, Appl  
169 7 2.9 565 14 US-10-065-133A-11 Sequence 11, Appl  
170 7 2.9 565 14 US-10-065-133A-8 Sequence 8, Appl  
171 7 2.9 565 16 US-10-434-811A-11 Sequence 11, Appl  
172 7 2.9 585 12 US-10-425-114-53676 Sequence 53676, A  
173 7 2.9 601 9 US-09-480-236-1 Sequence 1, Appl  
174 7 2.9 613 10 US-09-480-236-3 Sequence 3, Appl  
175 7 2.9 613 10 US-09-462-713-2 Sequence 2, Appl  
176 7 2.9 618 9 US-09-817-676A-14 Sequence 14, Appl  
177 7 2.9 618 9 US-09-970-516-4 Sequence 4, Appl  
178 7 2.9 618 14 US-10-354-358-78 Sequence 78, Appl  
179 7 2.9 622 9 US-09-147-346-4 Sequence 4, Appl  
180 7 2.9 676 11 US-09-873-155-40 Sequence 40, Appl  
181 7 2.9 676 14 US-10-096-840D-4 Sequence 4, Appl  
182 7 2.9 693 12 US-10-425-114-56454 Sequence 56454, A  
183 7 2.9 703 12 US-10-425-114-58817 Sequence 58817, A  
184 7 2.9 731 14 US-10-094-113-6 Sequence 6, Appl  
185 7 2.9 951 15 US-10-363-493-3330 Sequence 3330, Ap  
186 7 2.9 1050 12 US-10-282-122A-67554 Sequence 67554, A  
187 7 2.9 1759 15 US-10-369-493-7032 Sequence 7032, Ap  
188 7 2.9 3939 14 US-10-156-761-10434 Sequence 10434, A  
189 6 2.5 6 14 US-10-206-699-122 Sequence 122, App  
190 6 2.5 6 14 US-10-032-221B-48 Sequence 48, Appl  
191 6 2.5 6 14 US-10-032-221B-54 Sequence 54, Appl  
192 6 2.5 7 14 US-10-032-221B-58 Sequence 58, Appl  
193 6 2.5 8 10 US-09-839-996-7 Sequence 7, Appl  
194 6 2.5 8 10 US-09-839-996-8 Sequence 8, Appl  
195 6 2.5 8 14 US-10-080-505-53 Sequence 53, Appl  
196 6 2.5 8 14 US-10-080-505-54 Sequence 54, Appl  
197 6 2.5 9 9 US-09-780-053-452 Sequence 452, App  
198 6 2.5 9 13 US-10-066-151-82 Sequence 82, Appl  
199 6 2.5 9 13 US-10-035-688-4 Sequence 4, Appl  
200 6 2.5 9 14 US-10-032-221B-57 Sequence 57, Appl

## ALIGNMENTS

## RESULT 1

US-10-032-221B-10  
; Sequence 10, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREC  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 244  
; TYPE: PRT

ORGANISM: Homo sapiens  
US-10-032-221B-10

Query Match 100.0%; Score 244; DB 14; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.9e-233;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTIGSCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120  
DB 61 LGTIGSCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALASPGSCLE 180  
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALASPGSCLE 180

QY 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKIIISRCQVCM 240  
DB 181 EFRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKIIISRCQVCM 240

QY 241 KKRH 244  
DB 241 KKRH 244

RESULT 2  
US-10-032-221B-22  
; Sequence 22, Application US/10032221B  
; Publication No. US20030144881A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghubar  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032.221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 22  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-1 (Tumstatin N-53) (amino acids 54-244 of SEQ ID NO:10)  
US-10-032-221B-22

Query Match 78.3%; Score 191; DB 14; Length 191;  
Best Local Similarity 100.0%; Pred. No. 5.4e-181;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ORAHGQDGLTGSCCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGR 113  
DB 1 ORAHGQDGLTGSCCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGR 60

QY 114 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALA 173  
DB 61 ALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTGALA 120

QY 174 SPGSCLEEFPRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKII 233  
DB 121 SPGSCLEEFPRASPFLECHGRGTCNYNSYSYFWLASLNPFRMFRKPIPTVKGAGELEKII 180

QY 234 SRCQVCMKKRH 244  
DB 181 SRCQVCMKKRH 191

RESULT 3  
US-10-270-877-46  
; Sequence 46, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-877-46

Query Match 65.2%; Score 159; DB 14; Length 211;  
Best Local Similarity 100.0%; Pred. No. 3.1e-149;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGRGDSGSPATWTRGFTVTRHSQTATPSCPEGTVPYSGFSLFVQGNQRAHQD 60

QY 61 LGTIGSCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120  
DB 61 LGTIGSCLQRTTTPPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120

QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159  
DB 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159

RESULT 4  
US-10-270-837-46  
; Sequence 46, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDV  
US-10-270-837-46

US-10-270-837-46

Query Match 65.2%; Score 159; DB 14; Length 211;  
Best Local Similarity 100.0%; Pred. No. 3.1e-149;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGRGDSGPATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYIS 120  
DB 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAVHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 159  
DB 121 RCTVCEGPAIAVHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 159

RESULT 5

US-10-206-699-304  
; Sequence 304, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US/10/206,699  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 304  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 3 chain  
US-10-206-699-304

Query Match 63.5%; Score 155; DB 14; Length 232;  
Best Local Similarity 100.0%; Pred. No. 3.1e-145;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLQRT 72  
DB 1 ATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLQRT 60

QY 73 TWPFLFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYISRCTVCEGPAIAI 132  
DB 61 TWPFLFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYISRCTVCEGPAIAI 120

QY 133 AVHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLQRT 167  
DB 121 AVHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQDGLTGLSCLQRT 155

RESULT 6

US-10-032-221b-23  
; Sequence 23, Application US/10032221b  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Tum-2 (amino acids 1-132 of SEQ ID NO:10)  
US-10-032-221B-23

Query Match 54.1%; Score 132; DB 14; Length 132;  
Best Local Similarity 100.0%; Pred. No. 1.2e-122;  
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGPATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGRGDSGPATWTRGFTVTRHSQTATPSCPEGTVPVLYSGFSLFVQGNQRAHQD 60

QY 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYIS 120  
DB 61 LGTLGSCLOQRTTTFPFCNVNDVCFASRNDYSYWLSTPALMPMNNAPITGRALEPYIS 120

QY 121 RCTVCEGPAIAI 132  
DB 121 RCTVCEGPAIAI 132

RESULT 7

US-10-032-221B-20  
; Sequence 20, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 124  
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-20

Query Match      50.8%; Score 124; DB 14; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.8e-115;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKKGDSGSPATWTRGVFTRHSQTTAIPSCPEGTVPLVSGSFLEVQGNQRAHGD 60
   |||||
Db 1 GLKKGDSGSPATWTRGVFTRHSQTTAIPSCPEGTVPLVSGSFLEVQGNQRAHGD 60
   |||||

QY 61 LGTLGSLQRFTHMPFLFCNVNDVCFNPNFASRNDYSYMLSTPALPMNMAITGRALEPYIS 120
   |||||
Db 61 LGTLGSLQRFTHMPFLFCNVNDVCFNPNFASRNDYSYMLSTPALPMNMAITGRALEPYIS 120
   |||||

QY 121 RCTV 124
   |||||
Db 121 RCTV 124
   |||||

RESULT 8
US-10-032-221B-21
; Sequence 21, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 334 (amino acids 125-244 of SEQ ID NO:10)
US-10-032-221B-21

Query Match      49.2%; Score 120; DB 14; Length 120;
Best Local Similarity 100.0%; Pred. No. 8.8e-111;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 CEGPATAIVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 184
   |||||
Db 1 CEGPATAIVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRA 60
   |||||

QY 185 SPFLECHGRGTCNYNSYNSFWLASLNPERMFRKPIPTVKGAELEKIISRCQVCMKKRH 244
   |||||
Db 61 SPFLECHGRGTCNYNSYNSFWLASLNPERMFRKPIPTVKGAELEKIISRCQVCMKKRH 120
   |||||

RESULT 9
US-10-032-221B-24
; Sequence 24, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tum-3 (amino acids 133-244 of SEQ ID NO:10)
US-10-032-221B-24

Query Match      45.9%; Score 112; DB 14; Length 112;
Best Local Similarity 100.0%; Pred. No. 7.1e-103;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRAFPFLCHG 192
   |||||
Db 1 AVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEGTQALASPGSCLEEFRAFPFLCHG 60
   |||||

QY 193 RGTCTNYNSYNSFWLASLNPERMFRKPIPTVKGAELEKIISRCQVCMKKRH 244
   |||||
Db 61 RGTCTNYNSYNSFWLASLNPERMFRKPIPTVKGAELEKIISRCQVCMKKRH 112
   |||||

RESULT 10
US-10-032-221B-33
; Sequence 33, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin 333 (amino acids 1-124 of SEQ ID NO:10)
US-10-032-221B-33
```

```

;
; FEATURE:
; OTHER INFORMATION: Tumstatin-45-132 (amino acids 45-132 of SEQ ID NO:10)
US-10-032-221B-33

Query Match 36.1%; Score 88; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 FSEFLVQGNQRAHGQDLGTGSCLCQRFRTTNPFFLCNNVNDVCNFSASNDYSYWLSTPALMP 104
Db 1 FSEFLVQGNQRAHGQDLGTGSCLCQRFRTTNPFFLCNNVNDVCNFSASNDYSYWLSTPALMP 60

QY 105 MNMAPITGRALEPYISRCTVCEGPAIAI 132
Db 61 MNMAPITGRALEPYISRCTVCEGPAIAI 88

RESULT 11
US-10-032-221B-34
; Sequence 34, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 34
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumstatin-5-125-C-A (amino acids 45-132 of SEQ ID NO:10; alanine
; OTHER INFORMATION: has been substituted for the cysteine residue at position 125 of
; OTHER INFORMATION: the full-length Tumstatin molecule)
US-10-032-221B-34

Query Match 32.8%; Score 80; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 3e-71;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 FSEFLVQGNQRAHGQDLGTGSCLCQRFRTTNPFFLCNNVNDVCNFSASNDYSYWLSTPALMP 104
Db 1 FSEFLVQGNQRAHGQDLGTGSCLCQRFRTTNPFFLCNNVNDVCNFSASNDYSYWLSTPALMP 60

QY 105 MNMAPITGRALEPYISRCTV 124
Db 61 MNMAPITGRALEPYISRCTV 80

RESULT 12
US-10-032-221B-26
; Sequence 26, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF

```



Query Match 26.2%; Score 64; DB 14; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.7e-55;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 181 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRKPIPTSTVKAGELEKIIISRCQVCM 240  
DB 1 EFRASPFLECHGRGTCNYNSYSFWLASLNPFRKPIPTSTVKAGELEKIIISRCQVCM 60  
QY 241 KKRH 244  
DB 61 KKRH 64

RESULT 14  
US-10-270-877-50  
; Sequence 50, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 50  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V  
US-10-270-877-50

Query Match 25.0%; Score 61; DB 14; Length 68;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
QY 61 L 61  
DB 61 L 61

RESULT 15  
US-10-270-837-50  
; Sequence 50, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 50  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V

US-10-270-837-50

Query Match 25.0%; Score 61; DB 14; Length 68;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
QY 61 L 61  
DB 61 L 61

RESULT 16  
US-10-270-877-48  
; Sequence 48, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII  
US-10-270-877-48

Query Match 25.0%; Score 61; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
DB 1 GLKGKRGDSGPATWTTRGFVTRHSQTTAIPSCPEGTVPLYSGFSLFVQGNQRAHQD 60  
QY 61 L 61  
DB 61 L 61

RESULT 17  
US-10-270-877-52  
; Sequence 52, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 52  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V  
US-10-270-837-52

Query Match 25.0%; Score 61; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60  
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61  
DB 61 L 61

## RESULT 18

US-10-270-837-48  
; Sequence 48, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII

## US-10-270-837-48

Query Match 25.0%; Score 61; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60  
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61  
DB 61 L 61

## RESULT 19

US-10-270-837-52  
; Sequence 52, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 52  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

## US-10-270-837-52

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V  
US-10-270-837-52

Query Match 25.0%; Score 61; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52; Indels 0; Gaps 0;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60  
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 60

QY 61 L 61  
DB 61 L 61

## RESULT 20

US-10-270-877-61  
; Sequence 61, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII

## US-10-270-877-61

Query Match 15.2%; Score 37; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28; Indels 0; Gaps 0;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 37  
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTPSCPEGTVPVLYSGFSFLFVQGNORAHGQD 37

## RESULT 21

US-10-270-837-61  
; Sequence 61, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII

## US-10-270-837-61

Query Match 15.2%; Score 37; DB 14; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28; Indels 0; Gaps 0;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 266  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-266

Query Match 9.0%; Score 22; DB 14; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.8e-14;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTMPELCNVNDVCNFSRND 92  
DB 1 FTTMPELCNVNDVCNFSRND 22

## RESULT 26

US-10-270-877-26  
; Sequence 26, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:

; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: GPpepl

US-10-270-877-26

Query Match 8.6%; Score 21; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGKRGDSGSPATWTRGFVFT 23  
DB 1 KGKRGDSGSPATWTRGFVFT 21

## RESULT 27

US-10-270-837-26  
; Sequence 26, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:

; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11

; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl  
US-10-270-837-26

Query Match 8.6%; Score 21; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KGKRGDSGSPATWTRGFVFT 23  
DB 1 KGKRGDSGSPATWTRGFVFT 21

## RESULT 28

US-10-206-699-247  
; Sequence 247, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 247  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-247

Query Match 8.6%; Score 21; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 FWLASLNPERMFKRPISTVK 225  
DB 1 FWLASLNPERMFKRPISTVK 21

## RESULT 29

US-10-206-699-289  
; Sequence 289, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:

; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289

; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 289  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-693-289

Query Match 8.2%; Score 20; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTHMPFLFCNVNVCNF 87  
Db 1 LQRFTHMPFLFCNVNVCNF 20

## RESULT 30

US-10-032-221B-28  
; Sequence 28, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

; OTHER INFORMATION: T2 (amino acids 53-72 of SEQ ID NO:10)  
US-10-032-221B-28

Query Match 8.2%; Score 20; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 NORAHGQDLGTLGSCLOQRT 72  
Db 1 NORAHGQDLGTLGSCLOQRT 20

## RESULT 31

US-10-032-221B-29  
; Sequence 29, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)

; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T3 (amino acids 58-87 of SEQ ID NO:10)  
US-10-032-221B-29

Query Match 8.2%; Score 20; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LQRFTHMPFLFCNVNVCNF 87  
Db 1 LQRFTHMPFLFCNVNVCNF 20

## RESULT 32

US-10-032-221B-30  
; Sequence 30, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:

; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T4 (amino acids 83-102 of SEQ ID NO:10)  
US-10-032-221B-30

Query Match 8.2%; Score 20; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNFSRNDYSYWLSTPAL 102  
Db 1 DVCNFSRNDYSYWLSTPAL 20

## RESULT 33

US-10-032-221B-27  
; Sequence 27, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T1 (amino acids 1-19 of SEQ ID NO:10)  
US-10-032-221B-27

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRG 19  
Db 1 GLKGRGDSGSPATWTRG 19

## RESULT 34

US-10-032-221B-31  
; Sequence 31, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 31  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T5 (amino acids 98-116 of SEQ ID NO:10)  
US-10-032-221B-31

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 STPALPMNMNMAPITGRALE 116  
Db 1 STPALPMNMNMAPITGRALE 19

## RESULT 35

US-10-032-221B-32  
; Sequence 32, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 32  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T6 (amino acids 113-131 of SEQ ID NO:10)  
US-10-032-221B-32

Query Match 7.8%; Score 19; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.3e-11;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RALEPYISRCTVCEGPAIA 131  
Db 1 RALEPYISRCTVCEGPAIA 19

## RESULT 36

US-10-206-699-254  
; Sequence 254, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523

; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 254  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-254

Query Match 7.4%; Score 18; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 FTTFPFLFCNVNVCNFA 88  
|||||  
Db 1 FTTFPFLFCNVNVCNFA 18

## RESULT 37

US-10-206-699-260  
; Sequence 260, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 260  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-260

Query Match 7.4%; Score 18; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNVCNFA 92  
|||||  
Db 1 PFLFCNVNVCNFA 18

## RESULT 38

US-10-206-699-277  
; Sequence 277, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 277  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-277

Query Match 7.4%; Score 18; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 LNPFMRKPISTVKAG 227  
|||||  
Db 1 LNPFMRKPISTVKAG 18

## RESULT 39

US-09-864-761-48095  
; Sequence 48095, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David K.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29



; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 48095  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL035425.11  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
; OTHER INFORMATION: EST HUMAN HIT: AW893189.1, EVALUE 2.00e-22  
; OTHER INFORMATION: SWISSPROT HIT: P29400, EVALUE 2.00e-23  
US-09-864-761-48095

Query Match 7.0%; Score 17; DB 9; Length 46;  
Best Local Similarity 100.0%; Pred. No. 4.9e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100  
Db 10 VCNFASRNDYSYWLSTP 26  
|||||

## RESULT 40

US-10-206-699-302  
; Sequence 302, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MHBB 01-1017  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 302  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 1 chain  
US-10-206-699-302

Query Match 7.0%; Score 17; DB 14; Length 229;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100  
Db 70 VCNFASRNDYSYWLSTP 86  
|||||

## RESULT 41

US-10-206-699-306  
; Sequence 306, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MHBB 01-1017  
; CURRENT FILING DATE: 2002-07-26  
US-10-206-699-306

; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 306  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: alpha 5 chain  
US-10-206-699-306

Query Match 7.0%; Score 17; DB 14; Length 229;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100  
Db 70 VCNFASRNDYSYWLSTP 86  
|||||

## RESULT 42

US-10-032-221B-2  
; Sequence 2, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-2

Query Match 7.0%; Score 17; DB 14; Length 229;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASRNDYSYWLSTP 100  
Db 70 VCNFASRNDYSYWLSTP 86  
|||||

## RESULT 43

US-09-925-297-496  
; Sequence 496, Application US/09925297  
; Patent No. US20020081659A1

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (247)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-496

Query Match          7.0%; Score 17; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      150 VCNFASRNDYSYWLSTP 166
|||||

RESULT 44
US-09-925-302-507
; Sequence 507, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 507
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-507

Query Match          7.0%; Score 17; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      247 VCNFASRNDYSYWLSTP 263
|||||

RESULT 45
US-10-372-683-8
; Sequence 8, Application US/10372683
; Publication No. US20040009171A1
; GENERAL INFORMATION:
; APPLICANT: GERRITSEN, MARY E.
; APPLICANT: PEALE JR., FRANKLIN V.
```

```
; APPLICANT: WU, THOMAS D.
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA
; FILE REFERENCE: P1928R1P1
; CURRENT APPLICATION NUMBER: US/10/372,683
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 10/271,690
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/344,534
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 8
; LENGTH: 1669
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-372-683-8

Query Match          7.0%; Score 17; DB 15; Length 1669;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYSYWLSTP 100
Db      1510 VCNFASRNDYSYWLSTP 1526
|||||

RESULT 46
US-10-206-699-74
; Sequence 74, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-74

Query Match          6.8%; Score 16; DB 14; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      50 VQGNQRAHGQDLGTLG 65
Db      1 VQGNQRAHGQDLGTLG 16
|||||

RESULT 47
US-10-206-699-53
; Sequence 53, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
```

```
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-53

Query Match
Best Local Similarity 6.1%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLNPERMFRKPIPT 223
Db 1 SLNPERMFRKPIPT 15

RESULT 48
US-10-206-699-94
; Sequence 94, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 94
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-94

Query Match
Best Local Similarity 6.1%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TSAGSEGTGQALASP 175
Db 1 TSAGSEGTGQALASP 15

RESULT 49
US-10-206-699-137
; Sequence 137, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
```

```
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 137
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-137

Query Match
Best Local Similarity 6.1%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PALMPMNWAPITGRA 114
Db 1 PALMPMNWAPITGRA 15

RESULT 50
US-10-206-699-191
; Sequence 191, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-191

Query Match
Best Local Similarity 6.1%; Score 15; DB 14; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 FRKPIPTVKAGELE 230
Db 1 FRKPIPTVKAGELE 15

RESULT 51
US-10-206-699-212
; Sequence 212, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
```

; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 212  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-212

Query Match 6.1%; Score 15; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07; Length 15;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 DVCNPFASRNDYSYWL 97  
|||||  
DB 1 DVCNPFASRNDYSYWL 15

## RESULT 52

US-10-206-699-231  
; Sequence 231, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 231  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-231

Query Match 6.1%; Score 15; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07; Length 15;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 RGTGNYNSYSYFWL 207  
|||||  
DB 1 RGTGNYNSYSYFWL 15

## RESULT 53

US-10-206-699-3  
; Sequence 3, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-3

Query Match 5.7%; Score 14; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06; Length 14;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PFLFCNVNDVCNPA 88  
|||||  
DB 1 PFLFCNVNDVCNPA 14

## RESULT 54

US-10-206-699-28  
; Sequence 28, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-28

Query Match 5.7%; Score 14; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06; Length 14;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 SCLQRTTWPFLFC 79  
|||||  
DB 1 SCLQRTTWPFLFC 14

## RESULT 55

US-10-206-699-112  
; Sequence 112, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854

;; PRIOR FILING DATE: 2002-03-22  
;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 112  
;; LENGTH: 14  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-112

Query Match 5.7%; Score 14; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 SELFVQGNQRAHQ 59  
DB 1 SELFVQGNQRAHQ 14

RESULT 56  
US-10-206-699-165  
;; Sequence 165, Application US/10206699  
;; Publication No. US20030100510A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sundaramoorthy, M.  
;; APPLICANT: Hudson, B.  
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
;; FILE REFERENCE: MBHB 01-1017  
;; CURRENT APPLICATION NUMBER: US/10/206,699  
;; CURRENT FILING DATE: 2002-07-26  
;; PRIOR APPLICATION NUMBER: US 60/308,523  
;; PRIOR FILING DATE: 2001-07-27  
;; PRIOR APPLICATION NUMBER: US 60/351,289  
;; PRIOR FILING DATE: 2001-10-29  
;; PRIOR APPLICATION NUMBER: US 60/366,854  
;; PRIOR FILING DATE: 2002-03-22  
;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 165  
;; LENGTH: 14  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-165

Query Match 5.7%; Score 14; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGSEGTGQA 171  
DB 1 IMFTSAGSEGTGQA 14

RESULT 57  
US-10-206-699-272  
;; Sequence 272, Application US/10206699  
;; Publication No. US20030100510A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sundaramoorthy, M.  
;; APPLICANT: Hudson, B.  
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
;; FILE REFERENCE: MBHB 01-1017  
;; CURRENT APPLICATION NUMBER: US/10/206,699  
;; CURRENT FILING DATE: 2002-07-26  
;; PRIOR APPLICATION NUMBER: US 60/308,523  
;; PRIOR FILING DATE: 2001-07-27  
;; PRIOR APPLICATION NUMBER: US 60/351,289  
;; PRIOR FILING DATE: 2001-10-29  
;; PRIOR APPLICATION NUMBER: US 60/366,854  
;; PRIOR FILING DATE: 2002-03-22

;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 272  
;; LENGTH: 14  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-272

Query Match 5.7%; Score 14; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 PFLECHGRGTCNYY 199  
DB 1 PFLECHGRGTCNYY 14

RESULT 58  
US-10-206-699-210  
;; Sequence 210, Application US/10206699  
;; Publication No. US20030100510A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sundaramoorthy, M.  
;; APPLICANT: Hudson, B.  
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
;; FILE REFERENCE: MBHB 01-1017  
;; CURRENT APPLICATION NUMBER: US/10/206,699  
;; CURRENT FILING DATE: 2002-07-26  
;; PRIOR APPLICATION NUMBER: US 60/308,523  
;; PRIOR FILING DATE: 2001-07-27  
;; PRIOR APPLICATION NUMBER: US 60/351,289  
;; PRIOR FILING DATE: 2001-10-29  
;; PRIOR APPLICATION NUMBER: US 60/366,854  
;; PRIOR FILING DATE: 2002-03-22  
;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 210  
;; LENGTH: 15  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-210

Query Match 5.7%; Score 14; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWL 97  
DB 2 VCNFASRNDYSYWL 15

RESULT 59  
US-10-270-877-27  
;; Sequence 27, Application US/10270877  
;; Publication No. US20030049791A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Saus, Juan  
;; TITLE OF INVENTION: Goodpasture Binding Protein  
;; FILE REFERENCE: 98-723-AD1  
;; CURRENT APPLICATION NUMBER: US/10/270,877  
;; CURRENT FILING DATE: 2002-10-11  
;; PRIOR APPLICATION NUMBER: 09/512,563  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: 60/121,483  
;; PRIOR FILING DATE: 1993-02-24  
;; NUMBER OF SEQ ID NOS: 63  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 27  
;; LENGTH: 21

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpep1a9
US-10-270-877-27
```

```
Query Match          5.7%; Score 14; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3e-06; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 GSPATWTRGFVFT 23
      |||||
DB 8 GSPATWTRGFVFT 21
```

## RESULT 60

```
US-10-270-837-27
; Sequence 27, Application US/10270837
; Publication No. US20030054488A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-AD2
; CURRENT APPLICATION NUMBER: US/10/270,837
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 09/512,563
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: GPpep1a9
US-10-270-837-27
```

```
Query Match          5.7%; Score 14; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3e-06; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 GSPATWTRGFVFT 23
      |||||
DB 8 GSPATWTRGFVFT 21
```

## RESULT 61

```
US-10-032-221B-38
; Sequence 38, Application US/10032221B
; Publication No. US20030144481A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Kalluri, Raghuram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
```

```
; SOFTWARE: Patent in version 3.1
```

```
; SEQ ID NO 38
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: T7-mutant (amino acids 73-97 of SEQ ID NO:10; methionine has been substituted for the leucine residue at position 77 of the full-length Tumor necrosis factor molecule, and isoleucine has been substituted for the valine at position 81, and asparagine has been substituted for the aspartic acid at position 83)
US-10-032-221B-38
```

```
Query Match          5.7%; Score 14; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e-06; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 84 VCNFASRNDYSYWL 97
      |||||
DB 12 VCNFASRNDYSYWL 25
```

## RESULT 62

```
US-10-206-699-10
; Sequence 10, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-10
```

```
Query Match          4.9%; Score 12; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00014; Indels 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 186 PFLECHGRGTCN 197
      |||||
DB 1 PFLECHGRGTCN 12
```

## RESULT 63

```
US-10-206-699-150
; Sequence 150, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MEHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
```

; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 150  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-150

Query Match 4.5%; Score 11; DB 14; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LMPNMNAPITG 112  
|||||  
Db 1 LMPNMNAPITG 11

## RESULT 64

US-10-206-699-176  
; Sequence 176, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 176  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-176

Query Match 4.5%; Score 11; DB 14; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYY 199  
|||||  
Db 1 ECHGRGTCNYY 11

## RESULT 65

US-10-206-699-239  
; Sequence 239, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 239  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-239

Query Match 4.5%; Score 11; DB 14; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 LNPFRFRKPI 220  
|||||  
Db 1 LNPFRFRKPI 11

## RESULT 66

US-10-206-699-271  
; Sequence 271, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 271  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-271

Query Match 4.5%; Score 11; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ECHGRGTCNYY 199  
|||||  
Db 4 ECHGRGTCNYY 14

## RESULT 67

US-10-032-221B-41  
; Sequence 41, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCI/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118



```
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: TP3 (amino acids 76-94 of SEQ ID NO:10; lysine has been substituted
; OTHER INFORMATION: ed for the phenylalanine residue at position 76 of the full-length
; OTHER INFORMATION: h Tumstatin molecule, and cysteine has been substituted for the
; OTHER INFORMATION: aspartic acid at position 83)
US-10-032-221B-41
Query Match          4.5%; Score 11; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      84 VCNFASRNDYS 94
Db      9 VCNFASRNDYS 19

RESULT 68
US-10-206-699-14
; Sequence 14, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US 10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(5)
; OTHER INFORMATION: Amino acids at positions 1-5 are optionally absent, such that if
; OTHER INFORMATION: 5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(22)
; OTHER INFORMATION: Amino acids at positions 18-22 are optionally absent, such that
; OTHER INFORMATION: 18 is absent, 19-22 are absent, if 19 is absent, 20-22 are abse
US-10-206-699-15
Query Match          4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      189 ECHGRGTCNYY 199
Db      9 ECHGRGTCNYY 19

RESULT 70
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832
Query Match          4.5%; Score 11; DB 15; Length 1744;
```

```
RESULT 69
US-10-206-699-16
; Sequence 16, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US 10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(5)
; OTHER INFORMATION: Amino acids at positions 1-5 are optionally absent, such that if
; OTHER INFORMATION: 5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(22)
; OTHER INFORMATION: Amino acids at positions 18-22 are optionally absent, such that
; OTHER INFORMATION: 18 is absent, 19-22 are absent, if 19 is absent, 20-22 are abse
US-10-206-699-16
Query Match          4.5%; Score 11; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      189 ECHGRGTCNYY 199
Db      9 ECHGRGTCNYY 19

RESULT 70
US-10-369-493-5832
; Sequence 5832, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5832
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5832
Query Match          4.5%; Score 11; DB 15; Length 1744;
```

Best Local Similarity 100.0%; Pred. No. 0.11; DB 14; Length 10;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLEEFRA 184  
| | | | | | | | | |  
DB 1675 SPGSCLEEFRA 1685

## RESULT 71

US-09-572-404B-132  
; Sequence 132, Application US/09572404B  
; Publication No. US20030078374A1  
; GENERAL INFORMATION:  
; APPLICANT: Proteom Ltd  
; TITLE OF INVENTION: Complementary peptide ligands from the human genome  
; FILE REFERENCE: Human patent  
; CURRENT APPLICATION NUMBER: US/09/572,404B  
; CURRENT FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 4203  
; SOFTWARE: ProtPatent version 1.0  
; SEQ ID NO 132  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; OTHER INFORMATION: sequence located in COL11A2 at 1584-1593 and may interact with Se  
; OTHER INFORMATION: 131 in this patent.  
US-09-572-404B-132

Query Match 4.1%; Score 10; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGSEG 167  
| | | | | | | | | |  
DB 1 IMFTSAGSEG 10

## RESULT 72

US-10-206-699-283  
; Sequence 283, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 283  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-283

Query Match 4.1%; Score 10; DB 14; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSCLQRFTTM 74  
| | | | | | | | | |  
DB 1 GSCLQRFTTM 10

## RESULT 73

US-10-206-699-298  
; Sequence 298, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 298  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-298

Query Match 4.1%; Score 10; DB 14; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FCNVNDVCNF 87  
| | | | | | | | | |  
DB 1 FCNVNDVCNF 10

## RESULT 74

US-10-032-221B-58  
; Sequence 58, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREX  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Generic Peptide  
US-10-032-221B-58

Query Match 4.1%; Score 10; DB 14; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYSYWL 97  
| | | | | | | | | |  
Db 1 ASRNDYSYWL 10

## RESULT 75

US-10-032-221B-40  
; Sequence 40, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 40  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T8-3 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8)  
; OTHER INFORMATION: ted for the leucine residue at position 68 of the full-length T8  
; OTHER INFORMATION: statin molecule, and serine has been substituted for the cysteine  
; OTHER INFORMATION: residues at positions 79 and 85)  
US-10-032-221B-40

Query Match 4.1%; Score 10; DB 14; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.027; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMPFLF 78  
| | | | | | | | | |  
Db 2 QRFTTMPFLF 11

## RESULT 76

US-10-032-221B-42  
; Sequence 42, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17

; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: P2 (amino acids 68-94 of SEQ ID NO:10; lysine has been substituted for the leucine residue at position 68 of the full-length T8)  
; OTHER INFORMATION: tatin molecule, and aspartic acid has been substituted for the  
; OTHER INFORMATION: steine residues at positions 79 and 85)  
US-10-032-221B-42

Query Match 4.1%; Score 10; DB 14; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.027; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMPFLF 78  
| | | | | | | | | |  
Db 2 QRFTTMPFLF 11

## RESULT 77

US-10-043-487-342  
; Sequence 342, Application US/10043487  
; Publication No. US20030055220A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, LEGRAIN  
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypept  
; TITLE OF INVENTION: mammalian polypeptides  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 342  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Shigella Flexneri  
US-10-043-487-342

Query Match 4.1%; Score 10; DB 14; Length 143;

Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
| | | | | | | | | |  
Db 27 PAIAIAVHSQ 36

## RESULT 78

US-10-206-699-303  
; Sequence 303, Application US/10206699  
; Publication No. US2003010051CA1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854

; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 303  
; LENGTH: 227  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: alpha 2 chain  
US-10-206-699-303

Query Match 4.1%; Score 10; DB 14; Length 227;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
Db 111 PAIAIAVHSQ 120

## RESULT 79

US-10-032-221B-6  
; Sequence 6, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Rachuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 227  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-032-221B-6.

Query Match 4.1%; Score 10; DB 14; Length 227;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
Db 111 PAIAIAVHSQ 120

## RESULT 80

US-09-925-302-518  
; Sequence 518, Application US/09925302  
; Patent No. US2002004941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302

; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 518  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (11)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-518

Query Match 4.1%; Score 10; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
Db 314 PAIAIAVHSQ 323

## RESULT 81

US-10-331-496A-27  
; Sequence 27, Application US/10331496A  
; Publication No. US20030228305A1  
; GENERAL INFORMATION:  
; APPLICANT: FRANTZ, GRETCHEN  
; APPLICANT: HILLAN, KENNETH J.  
; APPLICANT: PHILLIPS, HEIDI S.  
; APPLICANT: POLAKIS, PAUL  
; APPLICANT: SMITH, VICTORIA  
; APPLICANT: SPENCER, SUSAN D.  
; APPLICANT: WILLIAMS, P. MICKEY  
; APPLICANT: WU, THOMAS D.  
; APPLICANT: ZHANG, ZHEN  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF TUMOR  
; FILE REFERENCE: PS01491-PCT  
; CURRENT APPLICATION NUMBER: US/10/331,496A  
; CURRENT FILING DATE: 2002-12-30  
; PRIOR APPLICATION NUMBER: US 60/345,444  
; PRIOR FILING DATE: 2002-01-02  
; PRIOR APPLICATION NUMBER: US 60/351,885  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: US 60/360,066  
; PRIOR FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: US 60/362,004  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/366,869  
; PRIOR FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: US 60/366,284  
; PRIOR FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 60/368,679  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 60/404,809  
; PRIOR FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: US 60/405,645  
; PRIOR FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 95  
; SEQ ID NO 27  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-331-496A-27

Query Match 4.1%; Score 10; DB 15; Length 459;  
Best Local Similarity 100.0%; Pred. No. 0.32;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 PAIAIAVHSQ 137  
|||||  
Db 343 PAIAIAVHSQ 352

## RESULT 82

US-10-372-683-30  
; Sequence 30, Application US/10372683  
; Publication No. US20040009171A1  
; GENERAL INFORMATION:  
; APPLICANT: GERRITSEN, MARY E.  
; APPLICANT: PEALE JR., FRANKLIN V.  
; APPLICANT: WU, THOMAS D.  
; TITLE OF INVENTION: METHODS FOR THE TREATMENT OF CARCINOMA  
; FILE REFERENCE: P1928R1P1  
; CURRENT APPLICATION NUMBER: US/10/372,683  
; CURRENT FILING DATE: 2003-02-21  
; PRIOR APPLICATION NUMBER: US 10/271,690  
; PRIOR FILING DATE: 2002-10-16  
; PRIOR APPLICATION NUMBER: US 60/344,534  
; PRIOR FILING DATE: 2001-10-18  
; NUMBER OF SEQ ID NOS: 49  
; SEQ ID NO 30  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-372-683-30

Query Match 4.1%; Score 10; DB 15; Length 459;  
Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 128 PAIAIAVHSQ 137  
|||||  
Db 343 PAIAIAVHSQ 352

## RESULT 83

US-09-961-403-9  
; Sequence 9, Application US/09961403  
; Publication No. US20030077589A1  
; GENERAL INFORMATION:  
; APPLICANT: HE-STUMPP, HOLGER  
; APPLICANT: HAENDLER, BERNARD  
; APPLICANT: KRAETZSCHMAR, JOERN  
; APPLICANT: KREFT, BERTHOLT  
; APPLICANT: WINTERHAGER, ELKE  
; APPLICANT: REGIDOR, PEDRO  
; APPLICANT: SCOTTI, SIMONE  
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS  
; FILE REFERENCE: SCH-1789  
; CURRENT APPLICATION NUMBER: US/09/961,403  
; CURRENT FILING DATE: 2001-09-25  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1712  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-961-403-9

Query Match 4.1%; Score 10; DB 10; Length 1712;  
Best Local Similarity 100.0%; Pred. No. 1; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 128 PAIAIAVHSQ 137  
|||||  
Db 1596 PAIAIAVHSQ 1605

## RESULT 84

US-10-206-699-9

; Sequence 9, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-9

Query Match 3.7%; Score 9; DB 14; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.13; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 189 ECHGRGTCN 197  
|||||  
Db 4 ECHGRGTCN 12

## RESULT 85

US-10-206-699-76  
; Sequence 76, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 76  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-76

Query Match 3.7%; Score 9; DB 14; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 55 RAHQDGLGT 63  
|||||  
Db 6 RAHQDGLGT 14

## RESULT 86

US-10-206-699-259

; Sequence 259, Application US/10206699

Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 259  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-259

Query Match 3.7%; Score 9; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92  
| | | | | | | |  
DB 10 VCNFASRND 18

RESULT 87  
US-10-206-699-261  
; Sequence 261, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 261  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-261

Query Match 3.7%; Score 9; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92  
| | | | | | | |  
DB 10 VCNFASRND 18

RESULT 88  
US-10-206-699-265  
; Sequence 265, Application US/10206699  
; Publication No. US20030100510A1

GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 265  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-265

Query Match 3.7%; Score 9; DB 14; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92  
| | | | | | | |  
DB 14 VCNFASRND 22

RESULT 89  
US-10-206-699-267  
; Sequence 267, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 267  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-267

Query Match 3.7%; Score 9; DB 14; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRND 92  
| | | | | | | |  
DB 14 VCNFASRND 22

RESULT 90  
US-10-032-221B-50  
; Sequence 50, Application US/1003222B  
; Publication No. US2003014481A1  
; GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032-221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 50  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Generic Peptide  
US-10-032-221B-50

Query Match 3.3%; Score 8; DB 14; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.7e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LQFTTTP 75  
Db 1 LQFTTTP 8

RESULT 91  
US-10-032-221B-56  
; Sequence 56, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032-221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Generic Peptide  
US-10-032-221B-56

Query Match 3.3%; Score 8; DB 14; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.7e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 88 ASNDYSY 95  
Db 1 ASNDYSY 8  
RESULT 92  
US-10-270-877-44  
; Sequence 44, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V  
; OTHER INFORMATION: derived peptide  
US-10-270-877-44

Query Match 3.3%; Score 8; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61  
Db 1 QRAHGQDL 8

RESULT 93  
US-10-270-837-44  
; Sequence 44, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 09/512,563  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V  
; OTHER INFORMATION: derived peptide  
US-10-270-837-44

Query Match 3.3%; Score 8; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61



Db 1 QRAHGQDL 8

## RESULT 94

US-10-206-699-253  
; Sequence 253, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; APPLICANT: Hudson, B.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR FILING DATE: 2002-07-26  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 253  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-253

Query Match 3.3%; Score 8; DB 14; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMPFLFCN 80

Db 3 TMPFLFCN 10

## RESULT 95

US-10-270-877-43  
; Sequence 43, Application US/10270877  
; Publication No. US20030049791A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD1  
; CURRENT APPLICATION NUMBER: US/10/270,877  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR FILING DATE: 2002-10-11  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived  
US-10-270-877-43

Query Match 3.3%; Score 8; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61

Db 1 QRAHGQDL 8

## RESULT 96

US-10-270-837-43  
; Sequence 43, Application US/10270837  
; Publication No. US20030054488A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-AD2  
; CURRENT APPLICATION NUMBER: US/10/270,837  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived  
US-10-270-837-43

Query Match 3.3%; Score 8; DB 14; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHGQDL 61

Db 1 QRAHGQDL 8

## RESULT 97

US-10-206-699-288  
; Sequence 288, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
; FILE REFERENCE: MBHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 288  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-288

Query Match 3.3%; Score 8; DB 14; Length 20;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMPFLFCN 80

Db 6 TMPFLFCN 13

## RESULT 98

US-10-206-699-307  
; Sequence 307, Application US/10206699

Publication No. US20030100510A1  
GENERAL INFORMATION:  
APPLICANT: Sundaramoorthy, M.  
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NCI Domain Hexamer  
FILE REFERENCE: MBHB 01-1017  
CURRENT FILING DATE: 2002-07-26  
PRIOR APPLICATION NUMBER: US 60/308,523  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: US 60/351,289  
PRIOR FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: US 60/366,854  
PRIOR FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: US 60/385,362  
PRIOR FILING DATE: 2002-06-03  
NUMBER OF SEQ ID NOS: 307  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 307  
LENGTH: 228  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: alpha 6 chain  
US-10-206-699-307

Query Match 3.3%; Score 8; DB 14; Length 228;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCNK 241  
Db 219 SRCQVCNK 226

RESULT 99

US-10-369-493-3337  
Sequence 3337, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 39-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3337

LENGTH: 937

TYPE: PRT

ORGANISM: Neurospora crassa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(937)

OTHER INFORMATION: unsure at all Xaa locations

US-10-369-493-3337

Query Match 3.3%; Score 8; DB 15; Length 937;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 IMFTSAGS 165  
Db 576 IMFTSAGS 583

RESULT 100  
US-10-032-221B-49  
Sequence 49, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:  
APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 49  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Generic peptide  
US-10-032-221B-49

Query Match 2.9%; Score 7; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 9.7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRFTTMP 75  
Db 1 QRFTTMP 7

RESULT 101

US-10-032-221B-55  
Sequence 55, Application US/10032221B  
Publication No. US20030144481A1  
GENERAL INFORMATION:

APPLICANT: Kalluri, Raghuram  
TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THERE  
FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
CURRENT APPLICATION NUMBER: US/10/032,221B  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US01/00565  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: US 09/625,191  
PRIOR FILING DATE: 2000-07-21  
PRIOR APPLICATION NUMBER: US 09/543,371  
PRIOR FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: US 09/479,118  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: US 09/335,224  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/126,175  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/089,689  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 55  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Generic Peptide  
US-10-032-221B-55

Query Match 2.9%; Score 7; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 9.7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ASRNDYS 94  
|||||||  
Db 1 ASRNDYS 7

## RESULT 102

US-10-032-221B-51  
; Sequence 51, Application US/10032221B  
; Publication No. US20030144481A1  
; GENERAL INFORMATION:  
; APPLICANT: Kalluri, Raghuram  
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)  
; CURRENT APPLICATION NUMBER: US/10/032,221B  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: PCT/US01/00565  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/625,191  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 09/543,371  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: US 09/479,118  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: US 09/335,224  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/126,175  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/089,689  
; PRIOR FILING DATE: 1998-06-17  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Generic Peptide  
US-10-032-221B-51

Query Match 2.9%; Score 7; DB 14; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 QRETTMP 75  
|||||||  
Db 2 QRETTMP 8

## RESULT 103

US-10-206-699-26  
; Sequence 26, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MEHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22

; PRIOR APPLICATION NUMBER: US 60/385,362

; PRIOR FILING DATE: 2002-06-03

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 26

; LENGTH: 14

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-206-699-26

## Query Match

2.9%; Score 7; DB 14; Length 14;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 TMPFLFC 79  
|||||||  
Db 8 TMPFLFC 14

## RESULT 104

US-10-206-699-92  
; Sequence 92, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MEHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/385,362  
; PRIOR FILING DATE: 2002-06-03  
; NUMBER OF SEQ ID NOS: 307  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 92  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-206-699-92

## Query Match

2.9%; Score 7; DB 14; Length 15;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GQALASP 175  
|||||||  
Db 9 GQALASP 15

## RESULT 105

US-10-206-699-230  
; Sequence 230, Application US/10206699  
; Publication No. US20030100510A1  
; GENERAL INFORMATION:  
; APPLICANT: Sundaramoorthy, M.  
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
; FILE REFERENCE: MEHB 01-1017  
; CURRENT APPLICATION NUMBER: US/10/206,699  
; CURRENT FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: US 60/308,523  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/351,289  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/366,854  
; PRIOR FILING DATE: 2002-03-22

;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 230  
;; LENGTH: 15  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-230

Query Match 2.9%; Score 7; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 RGTCTNY 199  
Db 1 RGTCTNY 7

## RESULT 106

US-10-206-699-232  
;; Sequence 232, Application US/10206699  
;; Publication No. US20030100510A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sundaramoorthy, M.  
;; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
;; FILE REFERENCE: MEHB 01-1017  
;; CURRENT APPLICATION NUMBER: US/10/206,699  
;; CURRENT FILING DATE: 2002-07-26  
;; PRIOR APPLICATION NUMBER: US 60/308,523  
;; PRIOR FILING DATE: 2001-07-27  
;; PRIOR APPLICATION NUMBER: US 60/351,289  
;; PRIOR FILING DATE: 2001-10-29  
;; PRIOR APPLICATION NUMBER: US 60/366,854  
;; PRIOR FILING DATE: 2002-03-22  
;; PRIOR APPLICATION NUMBER: US 60/385,362  
;; PRIOR FILING DATE: 2002-06-03  
;; NUMBER OF SEQ ID NOS: 307  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 232  
;; LENGTH: 15  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-206-699-232

Query Match 2.9%; Score 7; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 RGTCTNY 199  
Db 1 RGTCTNY 7

## RESULT 107

US-09-864-761-37448  
;; Sequence 37448, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366

;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.5  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 37448  
;; LENGTH: 70  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL035425.11  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2  
;; OTHER INFORMATION: SWISSPROT HIT: P29400, EVALUATE 9.00e-24  
;; OTHER INFORMATION: EST\_HUMAN HIT: W07655.1, EVALUATE 6.00e-23  
US-09-864-761-37448

Query Match 2.9%; Score 7; DB 9; Length 70;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 TRHSQTT 29  
Db 32 TRHSQTT 38

## RESULT 108

US-09-864-761-47938  
;; Sequence 47938, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```
FILE REFERENCE: Aconica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 47938
LENGTH: 70
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL035425.10
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5
OTHER INFORMATION: EST HUMAN HIT: W07655.1, EVALUATE 6.00e-23
OTHER INFORMATION: SWISSPROT HIT: P29400, EVALUATE 9.00e-24
US-09-864-761-47938

Query Match      2.9%; Score 7; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 TRHSQTT 29
Db 32 TRHSQTT 38

RESULT 109
US-10-424-599-243322
; Sequence 243322, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1521
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
```

```
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 243322
LENGTH: 78
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_61748C.1.pep
US-10-424-599-243322

Query Match      2.9%; Score 7; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 SCLEEF 183
Db 38 SCLEEF 44

RESULT 110
US-10-424-599-182069
; Sequence 182069, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182069
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_135421C.1.pep
US-10-424-599-182069

Query Match      2.9%; Score 7; DB 12; Length 84;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 GFSFLV 50
Db 44 GFSFLV 50

RESULT 111
US-09-764-877-1521
; Sequence 1521, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1521
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
```

Query Match 2.9%; Score 7; DB 9; Length 87;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GKRGSQ 10  
Db 24 GKRGSQ 30

## RESULT 112

US-10-242-515-1521  
; Sequence 1521, Application US/10242515  
; Publication No. US2004009488A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005C1  
; CURRENT APPLICATION NUMBER: US/10/242,515  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 09/764,877  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1521  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-242-515-1521

Query Match 2.9%; Score 7; DB 15; Length 87;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GKRGSQ 10  
Db 24 GKRGSQ 30

## RESULT 113

US-10-425-114-54455  
; Sequence 54455, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 54455

; LENGTH: 111

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZWFLMO17114F11\_FLI.pap

US-10-425-114-54455

## Query Match

Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 HGQDLGT 63  
Db 102 HGQDLGT 108

## RESULT 114

US-09-461-580A-18  
; Sequence 18, Application US/09461580A  
; Publication No. US20030207325A1  
; GENERAL INFORMATION:  
; APPLICANT: Guarente, Leonard  
; APPLICANT: Imai, Shin-ichiro  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH  
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE  
; TITLE OF INVENTION: LIFESPAN  
; FILE REFERENCE: 0050.1618-000  
; CURRENT APPLICATION NUMBER: US/09/461,580A  
; CURRENT FILING DATE: 1999-12-15  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-461-580A-18

Query Match 2.9%; Score 7; DB 11; Length 117;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 KPIPTV 224  
Db 110 KPIPTV 116

## RESULT 115

US-09-864-761-38021  
; Sequence 38021, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38021  
LENGTH: 142  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031177.1  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.99  
OTHER INFORMATION: SWISSPROT HIT: Q14031, EVALUE 3.00e-55  
OTHER INFORMATION: EST\_HUMAN HIT: AUL42039.1, EVALUE 2.00e-42  
US-09-864-761-38021

Query Match 2.9%; Score 7; DB 9; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 AIAVHSQ 137  
DB 82 AIAVHSQ 88

RESULT 116  
US-10-094-749-3238  
Sequence 3238, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YANAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIC, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHIKO  
APPLICANT: YOSHIKAWA, TSUTOMU

APPLICANT: OTSUKA, MOTOYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3238  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-3238

Query Match 2.9%; Score 7; DB 15; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
DB 55 SGSPATW 61

RESULT 117  
US-10-425-114-47774  
Sequence 47774, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovacic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 47774  
LENGTH: 174  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3960-011-H7\_FLI.pgp  
US-10-425-114-47774

Query Match 2.9%; Score 7; DB 12; Length 174;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
DB 68 ALASPGS 74

RESULT 118  
US-09-727-855B-7  
Sequence 7, Application US/09727855B  
Patent No. US20020168703A1  
GENERAL INFORMATION:  
APPLICANT: HOSHINO, Tatsuo  
APPLICANT: OJIMA, Kazuyuki  
APPLICANT: SETOGUCHI, Yutaka  
TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF CAROTENOIDS AND BIOLOGICALLY USE  
TITLE OF INVENTION: MATERIALS THEREOF  
FILE REFERENCE: C38435/111694



; CURRENT APPLICATION NUMBER: US/09/727,855B  
; CURRENT FILING DATE: 2000-12-01  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Phaffia rhodozyma  
US-09-727-855B-7

Query Match 2.9%; Score 7; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120  
Db 15 ALEPYIS 21

## RESULT 119

; Sequence 59362, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yinhua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 59362  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Oryza sativa nipponbare  
; FEATURE:

; OTHER INFORMATION: Clone ID: JC-OSLELIB3474052F08\_FLI.pep  
US-10-425-114-59362

Query Match 2.9%; Score 7; DB 12; Length 218;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120  
Db 63 ALEPYIS 69

## RESULT 120

; Sequence 96, Application US/10267682  
; Publication No. US2004003235A1  
; GENERAL INFORMATION:

; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.

; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TRANSMISSION

; NUMBER OF SEQUENCES: 239  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York

; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/267,682  
; FILING DATE: 08-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,223A  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cortuzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-029  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: unknown

; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 96:  
US-10-267-682-96

Query Match 2.9%; Score 7; DB 12; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTOGA 171  
Db 29 SEGTOGA 35

## RESULT 121

; Sequence 96, Application US/10267748  
; Publication No. US20040052820A1  
; GENERAL INFORMATION:

; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.

; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TRANSMISSION

; NUMBER OF SEQUENCES: 239  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/267,748  
; FILING DATE: 08-Oct-2002

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,223A  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 96:  
US-10-267-748-96

Query Match 2.9%; Score 7; DB 12; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGOA 171  
Db 29 SEGTGOA 35  
|||||

RESULT 122  
US-10-206-699-305  
Sequence 305, Application US/10206699  
Publication No. US20030100510A1  
GENERAL INFORMATION:  
APPLICANT: Sundaramoorthy, M.  
APPLICANT: Hudson, B.  
TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer  
FILE REFERENCE: MBHB 01-1017  
CURRENT APPLICATION NUMBER: US/10/206,699  
CURRENT FILING DATE: 2002-07-26  
PRIOR APPLICATION NUMBER: US 60/308,523  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: US 60/351,289  
PRIOR FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: US 60/366,854  
PRIOR FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: US 60/385,362  
PRIOR FILING DATE: 2002-06-03  
NUMBER OF SEQ ID NOS: 307  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 305  
LENGTH: 231  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: alpha 4 chain  
US-10-206-699-305

Query Match 2.9%; Score 7; DB 14; Length 231;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 SPQSCLE 180  
Db 159 SPQSCLE 165  
|||||

RESULT 123  
US-10-230-331-28  
Sequence 28, Application US/10230331

Publication No. US20030157513A1  
GENERAL INFORMATION:  
APPLICANT: RAJASEKHARAN, Ram  
TITLE OF INVENTION: A NOVEL TRIACYLGLYCEROL BIOSYNTHESIS IN THE CYTOSOL OF EUKARYOTE  
FILE REFERENCE: 110522  
CURRENT APPLICATION NUMBER: US/10/230,331  
CURRENT FILING DATE: 2002-08-29  
PRIOR APPLICATION NUMBER: US 60/315,757  
PRIOR FILING DATE: 2001-08-30  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 28  
LENGTH: 233  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-230-331-28

Query Match 2.9%; Score 7; DB 14; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ALEPYIS 120  
Db 39 ALEPYIS 45  
|||||

RESULT 124  
US-10-282-122A-72960  
Sequence 72960, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 72960  
LENGTH: 247  
TYPE: PRT

; ORGANISM: Salmonella paratyphi A  
US-10-282-122A-72960

Query Match 2.9%; Score 7; DB 12; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212  
|||||  
DB 157 WLASLNP 163

## RESULT 125

US-10-282-122A-76032  
; Sequence 76032, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-03-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 76032

; TYPE: PRT

; ORGANISM: Salmonella typhi

US-10-282-122A-76032

Query Match 2.9%; Score 7; DB 12; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 WLASLNP 212  
|||||  
DB 157 WLASLNP 163

## RESULT 126

US-10-296-115-808

; Sequence 808, Application US/10296115  
; Publication No. US20040053248A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq Inc

; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 784PCT

; CURRENT APPLICATION NUMBER: US/10/296,115

; CURRENT FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: US09/488,725

; PRIOR FILING DATE: 2000-01-21

; PRIOR APPLICATION NUMBER: US09/552,317

; PRIOR FILING DATE: 2000-04-25

; NUMBER OF SEQ ID NOS: 1478

; SEQ ID NO: 808

; LENGTH: 251

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(251)

; OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3

US-10-296-115-808

Query Match 2.9%; Score 7; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
|||||  
DB 185 ALASPGS 191

## RESULT 127

US-10-424-599-235900

; Sequence 235900, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO: 235900

; LENGTH: 300

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_55045C.1.pgp

US-10-424-599-235900

Query Match 2.9%; Score 7; DB 12; Length 300;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 LEEFRAS 185  
|||||  
DB 147 LEEFRAS 153

## RESULT 128

US-09-964-956-71

; Sequence 71, Application US/09964956

; Publication No. US20040043926A1

; GENERAL INFORMATION:

; APPLICANT: Gerlach, Valerie L

; APPLICANT: MacDougall, John R

; APPLICANT: Smithson, Glennda

; APPLICANT: Millet, Isabelle

```

; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: No. US20040043926a1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/276,667
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/294,823
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/304,868
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 71
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: CNH domain
; OTHER INFORMATION: Consensus Sequence
US-09-964-956-71

Query Match          2.9%; Score 7; DB 12; Length 304;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      229 LEKISR 235
Db      36 LEKISR 42

RESULT 129
US-10-425-114-59795
; Sequence 59795, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59795
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-095-C11_FLI.pep
US-10-425-114-59795

Query Match          2.9%; Score 7; DB 12; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 ALEPYIS 120
Db      128 ALEPYIS 134

RESULT 130
US-10-425-114-71839
; Sequence 71839, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71839
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73165F09_FLI.pep
US-10-425-114-71839

Query Match          2.9%; Score 7; DB 12; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 KRGDGSGS 11
Db      207 KRGDGSGS 213

RESULT 131
US-10-425-114-50147
; Sequence 50147, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B

```

```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59795
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-095-C11_FLI.pep
US-10-425-114-59795

Query Match          2.9%; Score 7; DB 12; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 ALEPYIS 120
Db      128 ALEPYIS 134

RESULT 130
US-10-425-114-71839
; Sequence 71839, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71839
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73165F09_FLI.pep
US-10-425-114-71839

Query Match          2.9%; Score 7; DB 12; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 KRGDGSGS 11
Db      207 KRGDGSGS 213

RESULT 131
US-10-425-114-50147
; Sequence 50147, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B

```

; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 50147  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700901538\_FLI.pep  
US-10-425-114-50147

Query Match 2.9%; Score 7; DB 12; Length 317;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 IPSTVKA 226  
|||  
DB 288 IPSTVKA 294

## RESULT 132

US-10-424-599-201843  
; Sequence 201843, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 201843  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(328)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_2428C.1.pep  
US-10-424-599-201843

Query Match 2.9%; Score 7; DB 12; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 IPSTVKA 226  
|||  
DB 299 IPSTVKA 305

## RESULT 133

US-09-918-568-58  
; Sequence 58, Application US/09918568  
; Patent No. US20020054882A1  
; GENERAL INFORMATION:  
; APPLICANT: Yoshitobu OKUNO et al.  
; TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING  
; TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., #800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/918,568  
; FILING DATE: 02-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/004,422  
; FILING DATE: January 8, 1998  
; APPLICATION NUMBER: 08/443,862  
; FILING DATE: May 22, 1995  
; APPLICATION NUMBER: 08/229,781  
; FILING DATE: April 19, 1994  
; APPLICATION NUMBER: 08/054,016  
; FILING DATE: April 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER: <Unknown>  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 347 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: <Unknown>  
; ANTI-SENSE: <Unknown>  
; FRAGMENT TYPE: <Unknown>  
; ORIGINAL SOURCE:  
; ORGANISM: <Unknown>  
; STRAIN: <Unknown>  
; INDIVIDUAL ISOLATE: <Unknown>  
; DEVELOPMENTAL STAGE: <Unknown>  
; HAPLOTYPE: <Unknown>  
; TISSUE TYPE: <Unknown>  
; CELL TYPE: <Unknown>  
; CELL LINE: <Unknown>  
; ORGANELLE: <Unknown>  
; IMMEDIATE SOURCE:  
; LIBRARY: <Unknown>  
; CLONE: <Unknown>  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: <Unknown>  
; MAP POSITION: <Unknown>  
; UNITS: <Unknown>  
; FEATURE:  
; NAME/KEY:  
; LOCATION:  
; IDENTIFICATION METHOD:  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS:  
; TITLE:  
; JOURNAL:  
; VOLUME:  
; ISSUE:  
; PAGES:  
; DATE:  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:  
US-09-918-568-58

Query Match

2.9%; Score 7; DB 9; Length 347;

```
Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 165 SEGTGQA 171
Db 155 SEGTGQA 161

RESULT 134
US-09-860-351-2
; Sequence 2, Application US/09860351
; Patent No. US20020077463A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 16105, A NOVEL PROTEIN HUMAN PHOSPHATASE
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 38155-20013.00
; CURRENT APPLICATION NUMBER: US/09/860,351
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/205,260
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-860-351-2

Query Match 2.9%; Score 7; DB 9; Length 352;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 135
US-10-353-690-4
; Sequence 44, Application US/10353690
; Publication No. US20030215840A1
; GENERAL INFORMATION:
; APPLICANT: Logan, Thomas Joseph
; APPLICANT: Chun, Miyoung
; APPLICANT: Galvin, Katherine M.
; APPLICANT: Healy, Aileen
; APPLICANT: Acton, Susan L.
; APPLICANT: Donoghue, Mary
; APPLICANT: Stagliano, Nancy
; APPLICANT: Perodin, Jacqueline
; APPLICANT: Rodrigue-Way, Amelie
; TITLE OF INVENTION: Methods and compositions for treating
; TITLE OF INVENTION: cardiovascular disease using 1682, 6169, 6193, 7771, 14395,
; TITLE OF INVENTION: 29002, 33216, 43726, 62922, 66156, 32427, 2402, 7747, 1720,
; TITLE OF INVENTION: 9151, 60491, 1371, 7077, 33207, 1419, 18036, 16105, 38650,
; TITLE OF INVENTION: 14245, 58848, 1870, 25856, 32394, 3484, 345, 9252, 9135,
; TITLE OF INVENTION: 10532, 18610, 8165, 2448, 2445, 64624, 84237, 8912, 2868,
; TITLE OF INVENTION: 283, 2554, 9464, 17799, 26686, 43848, 32135, 12208, 2914,
; TITLE OF INVENTION: 51130, 19489, 21833, 2917, 59590, 15992, 2094, 2252, 3474,
; TITLE OF INVENTION: 9792, 15400, 1452 or 6585 molecules
; FILE REFERENCE: MF102-018PIRNONNIM
; CURRENT APPLICATION NUMBER: US/10/353,690
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/353,224
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/364,529
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 60/373,861
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/376,287
; PRIOR FILING DATE: 2002-04-29

Query Match 2.9%; Score 7; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
Db 55 SGSPATW 61

RESULT 137
US-09-973-064-4
; Sequence 4, Application US/09973064
; Patent No. US20020106773A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
```

; TITLE OF INVENTION: Diseases  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,064  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-973-064-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15  
Db 55 SGSPATW 61

## RESULT 138

US-09-973-077-4  
; Sequence 4, Application US/09973077  
; Patent No. US20020114799A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,077  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-973-077-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15  
Db 55 SGSPATW 61

## RESULT 139

US-09-973-063-4  
; Sequence 4, Application US/09973063  
; Patent No. US2002011519A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,063  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4

; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-973-063-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15  
Db 55 SGSPATW 61

## RESULT 140

US-09-973-964-4  
; Sequence 4, Application US/09973964  
; Patent No. US20020115606A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,964  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: US 60/304,775  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-973-964-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGSPATW 15  
Db 55 SGSPATW 61

## RESULT 141

US-09-975-072-4  
; Sequence 4, Application US/09975072  
; Patent No. US20020115607A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/975,072  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-975-072-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
|||||  
Db 55 SGSPATW 61

## RESULT 142

US-09-972-038-4  
; Sequence 4, Application US/09972038  
; Patent No. US20020119155A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/972,038  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-972-038-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
|||||  
Db 55 SGSPATW 61

## RESULT 143

US-09-972-757-4  
; Sequence 4, Application US/09972757  
; Patent No. US20020119927A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/972,757  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-972-757-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
|||||  
Db 55 SGSPATW 61

## RESULT 144

US-09-973-965-4

; Sequence 4, Application US/09973965  
; Patent No. US20020124273A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,965  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: US 60/304,775  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-973-965-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
|||||  
Db 55 SGSPATW 61

## RESULT 145

US-09-973-941-4  
; Sequence 4, Application US/09973941  
; Patent No. US20020164655A1  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.  
; APPLICANT: Heichman, Karen  
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
; FILE REFERENCE: Protein Interactions in ND  
; CURRENT APPLICATION NUMBER: US/09/973,941  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: US 60/240,790  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: US 60/304,775  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-973-941-4

Query Match 2.9%; Score 7; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15  
|||||  
Db 55 SGSPATW 61

## RESULT 146

US-09-986-992-2  
; Sequence 2, Application US/09986992  
; Publication No. US20030027308A1  
; GENERAL INFORMATION:  
; APPLICANT: FLOWMAN, GREGORY D.  
; APPLICANT: WHYTE, DAVID

```
; APPLICANT: MANNING, GERARD
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN PHOSPHATASES IDENTIFIED FROM
; FILE REFERENCE: 038602/1277
; CURRENT APPLICATION NUMBER: US/09/986,992
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2000-11-13
; PRIOR FILING DATE: 2000-11-13
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-992-2

Query Match      2.9%; Score 7; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 147
US-09-971-782-4
; Sequence 4, Application US/09971782
; Publication No. US20030186317A1
; GENERAL INFORMATION:
; APPLICANT: Roch, Jean-Marc
; APPLICANT: Bartel, Paul L.
; APPLICANT: Heichman, Karen
; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative
; FILE REFERENCE: Protein Interactions in ND
; CURRENT APPLICATION NUMBER: US/09/971,782
; PRIOR FILING DATE: 2001-10-09
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-971-782-4

Query Match      2.9%; Score 7; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 148
US-09-971-782-4
; Sequence 4, Application US/09971782
; Publication No. US20030186317A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
```

```
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NACHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1699
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1699

Query Match      2.9%; Score 7; DB 15; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGSPATW 15
DB 55 SGSPATW 61

RESULT 149
US-10-311-764-1
; Sequence 1, Application US/10311764
; Publication No. US20040023245A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; AU-YOUNG, Janice K.
; APPLICANT: BAUGHN, Mariah R.; DING, Li
; APPLICANT: ELLIOTT, Vicki S.; GANDHI, Ameena R.
; APPLICANT: GRIFFIN, Jennifer A.; HAPALIA, April J.A.
; APPLICANT: KEARNEY, Liam; LEE, Ernestine A.
; APPLICANT: LU, Yan; NGUYEN, Daniel B.
; APPLICANT: ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
; APPLICANT: REDDY, Rooda M.; SANJANWALA, Madhusudan M.
; APPLICANT: STEWART, Elizabeth A.; TANG, Y. Tom
; APPLICANT: THORNTON, Michael B.; TRIBOULEY, Catherine M.
; APPLICANT: CHAWLA, Narinder K.; YANG, Junming
; APPLICANT: YAO, Monique G.; YUE, Henry
; TITLE OF INVENTION: PROTEIN PHOSPHATASES
; FILE REFERENCE: PI-0126 USN
; CURRENT APPLICATION NUMBER: US/10/311,764
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/US01/19442
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/212,447
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/213,746
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: US 60/215,210
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/216,529
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/218,080
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/220,117
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 372
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023245A1 8124196CD1
US-10-311-764-1

```

```

Query Match      2.9%; Score 7; DB 16; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 9 SGSPATW 15
   |||||
Db 55 SGSPATW 61

```

## RESULT 150

```

US-09-925-300-1655
; Sequence 1655, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen.
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: FA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patencin ver. 2.0
; SEQ ID NO 1855
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (325)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1655

```

```

Query Match      2.9%; Score 7; DB 9; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 9 SGSPATW 15
   |||||
Db 56 SGSPATW 62

```

```

Search completed: April 5, 2004, 07:44:40
Job time : 48 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 07:36:14 ; Search time 22 Seconds  
(without alignments)  
572.579 Million cell updates/sec

Title: US-10-032-221B-10  
Perfect score: 244  
Sequence: 1 GLKRGDSGSPATWTRGF.....KAGLEKLIISRCQVCMKKRH 244

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database : Issued Parents AA: \*  
1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep: \*  
2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep: \*  
3: /cgn2\_6/prodata/2/iaa/6A-COMB.pep: \*  
4: /cgn2\_6/prodata/2/iaa/6B-COMB.pep: \*  
5: /cgn2\_6/prodata/2/iaa/PCTUS-COMB.pep: \*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	66.8	268	4	US-09-589-927-6
2	163	66.8	268	4	US-09-277-665-6
3	163	66.8	268	4	US-09-589-987-6
4	159	65.2	211	4	US-09-512-563C-46
5	141	57.8	218	2	US-08-399-889-25
6	141	57.8	218	3	US-09-167-364-25
7	141	57.8	218	3	US-09-439-897-4
8	61	25.0	68	4	US-09-512-563C-50
9	61	25.0	72	4	US-09-512-563C-48
10	61	25.0	72	4	US-09-512-563C-52
11	39	16.0	471	2	US-08-399-889-24
12	39	16.0	471	3	US-09-167-364-24
13	39	16.0	471	3	US-09-439-897-2
14	37	15.2	72	4	US-09-512-563C-61
15	36	14.8	36	3	US-09-439-897-63
16	26	10.7	26	3	US-09-439-897-65
17	21	8.6	21	4	US-09-512-563C-26
18	17	7.0	260	4	US-09-589-927-2
19	17	7.0	260	4	US-09-277-665-2
20	17	7.0	260	4	US-09-589-987-2
21	17	7.0	264	4	US-09-589-927-10
22	17	7.0	264	4	US-09-277-665-10
23	15	6.1	15	3	US-09-589-987-10
24	15	6.1	15	3	US-09-439-897-53
25	15	6.1	15	3	US-09-439-897-57
26	15	6.1	15	3	US-09-439-897-59
27	15	6.1	15	3	US-09-439-897-61

101	7	2.9	2	US-08-453-848-21	Sequence 21, Appl	174	6	2.5	111	1	US-08-466-886-38	Sequence 38, Appl
102	7	2.9	3	US-09-169-027-15	Sequence 15, Appl	175	6	2.5	111	3	US-08-469-617-38	Sequence 38, Appl
103	7	2.9	3	US-09-169-027-15	Sequence 15, Appl	176	6	2.5	121	4	US-09-621-976-7342	Sequence 7342, App
104	7	2.9	599	US-08-463-163-3	Sequence 3, Appl	177	6	2.5	122	4	US-09-325-932A-157	Sequence 157, App
105	7	2.9	613	US-08-405-615-1	Sequence 1, Appl	178	6	2.5	126	4	US-09-134-000C-5187	Sequence 5187, App
106	7	2.9	613	US-08-461-234-1	Sequence 1, Appl	179	6	2.5	128	4	US-09-252-991A-21538	Sequence 21538, A
107	7	2.9	613	US-08-463-480-1	Sequence 1, Appl	180	6	2.5	129	4	US-09-543-681A-6111	Sequence 6111, App
108	7	2.9	613	US-09-479-479-2	Sequence 2, Appl	181	6	2.5	131	4	US-09-553-498-4	Sequence 4, Appl
109	7	2.9	614	US-09-297-851-2	Sequence 2, Appl	182	6	2.5	131	4	US-09-618-869-4	Sequence 4, Appl
110	7	2.9	614	US-08-225-224-1	Sequence 1, Appl	183	6	2.5	135	3	US-09-293-395-9	Sequence 9, Appl
111	7	2.9	614	US-08-722-258-1	Sequence 1, Appl	184	6	2.5	135	3	US-09-668-648-9	Sequence 9, Appl
112	7	2.9	614	PCT-US95-04468-1	Sequence 1, Appl	185	6	2.5	140	4	US-09-724-138-44	Sequence 44, Appl
113	7	2.9	618	US-09-970-516-4	Sequence 4, Appl	186	6	2.5	144	4	US-09-252-991A-20518	Sequence 20518, A
114	7	2.9	622	US-08-356-786-16	Sequence 16, Appl	187	6	2.5	146	4	US-08-252-991A-23680	Sequence 23680, A
115	7	2.9	635	US-09-046-992-2	Sequence 2, Appl	188	6	2.5	150	4	US-09-636-215-707	Sequence 707, App
116	7	2.9	637	US-08-235-838-14	Sequence 14, Appl	189	6	2.5	150	4	US-09-685-166A-707	Sequence 707, App
117	7	2.9	637	US-08-235-838-16	Sequence 16, Appl	190	6	2.5	151	3	US-09-188-930-276	Sequence 276, App
118	7	2.9	637	US-08-465-473B-14	Sequence 14, Appl	191	6	2.5	151	4	US-09-312-283C-276	Sequence 276, App
119	7	2.9	637	US-08-465-473B-16	Sequence 16, Appl	192	6	2.5	152	4	US-09-621-976-4169	Sequence 4169, App
120	7	2.9	638	US-09-047-148-2	Sequence 2, Appl	193	6	2.5	153	3	US-08-828-741B-11	Sequence 11, Appl
121	7	2.9	658	US-08-252-991A-19879	Sequence 19879, A	194	6	2.5	153	4	US-08-925-433-4	Sequence 4, Appl
122	7	2.9	676	US-08-398-590A-40	Sequence 40, Appl	195	6	2.5	155	4	US-09-160-567-11	Sequence 11, Appl
123	7	2.9	676	US-08-894-997-40	Sequence 40, Appl	196	6	2.5	155	4	US-09-553-132-4	Sequence 4, Appl
124	7	2.9	767	US-09-252-991A-19361	Sequence 19361, A	197	6	2.5	155	4	US-09-710-299-11	Sequence 11, Appl
125	7	2.9	1694	US-08-494-168-2	Sequence 2, Appl	198	6	2.5	155	4	US-09-509-031-11	Sequence 11, Appl
126	7	2.9	3562	US-09-679-279-14	Sequence 14, Appl	199	6	2.5	166	2	US-08-729-103-4	Sequence 4, Appl
127	6	2.5	6	US-09-439-897-51	Sequence 51, Appl	200	6	2.5	168	2	US-08-702-105A-29	Sequence 29, Appl
128	6	2.5	7	US-10-080-505-58	Sequence 58, Appl							
129	6	2.5	8	US-08-296-791-7	Sequence 7, Appl							
130	6	2.5	8	US-08-296-791-8	Sequence 8, Appl							
131	6	2.5	8	US-09-839-996-7	Sequence 7, Appl							
132	6	2.5	8	US-09-839-996-8	Sequence 8, Appl							
133	6	2.5	8	US-10-080-505-53	Sequence 53, Appl							
134	6	2.5	8	US-10-080-505-54	Sequence 54, Appl							
135	6	2.5	8	PCT-US95-10661A-7	Sequence 7, Appl							
136	6	2.5	8	PCT-US95-10661A-8	Sequence 8, Appl							
137	6	2.5	9	US-09-322-624-4	Sequence 4, Appl							
138	6	2.5	11	US-10-080-505-20	Sequence 20, Appl							
139	6	2.5	12	US-08-260-582-40	Sequence 40, Appl							
140	6	2.5	12	PCT-US95-05471-40	Sequence 40, Appl							
141	6	2.5	15	US-09-439-897-58	Sequence 58, Appl							
142	6	2.5	16	US-08-106-568B-149	Sequence 149, App							
143	6	2.5	19	US-08-716-249-1	Sequence 1, Appl							
144	6	2.5	20	US-08-716-249-9	Sequence 9, Appl							
145	6	2.5	21	US-08-449-287-20	Sequence 20, Appl							
146	6	2.5	21	US-09-003-081-8	Sequence 8, Appl							
147	6	2.5	21	US-08-648-262-8	Sequence 8, Appl							
148	6	2.5	21	US-08-840-713-48	Sequence 48, Appl							
149	6	2.5	21	US-09-904-196B-3	Sequence 3, Appl							
150	6	2.5	21	US-09-230-233A-2	Sequence 2, Appl							
151	6	2.5	21	US-09-760-008A-3	Sequence 3, Appl							
152	6	2.5	21	US-09-393-171-19	Sequence 19, Appl							
153	6	2.5	21	US-08-318-193-46	Sequence 46, Appl							
154	6	2.5	22	US-08-215-138-1	Sequence 1, Appl							
155	6	2.5	23	US-08-407-344-1	Sequence 1, Appl							
156	6	2.5	26	US-08-666-354A-11	Sequence 11, Appl							
157	6	2.5	34	US-08-190-029A-6	Sequence 6, Appl							
158	6	2.5	34	US-08-462-695-6	Sequence 6, Appl							
159	6	2.5	35	US-08-318-193-47	Sequence 47, Appl							
160	6	2.5	62	US-09-621-976-5709	Sequence 5709, App							
161	6	2.5	62	US-09-621-976-5715	Sequence 5715, App							
162	6	2.5	68	US-09-134-001C-3669	Sequence 3669, App							
163	6	2.5	68	US-09-134-001C-3794	Sequence 3794, App							
164	6	2.5	68	US-09-134-001C-4489	Sequence 4489, App							
165	6	2.5	68	US-09-134-001C-5041	Sequence 5041, App							
166	6	2.5	72	US-09-134-001C-5281	Sequence 5281, App							
167	6	2.5	78	5187153-25	Patent No. 5187153							
168	6	2.5	78	5223482-27	Patent No. 5223482							
169	6	2.5	92	517841-1	Patent No. 517841							
170	6	2.5	96	US-09-390-134B-53	Sequence 53, Appl							
171	6	2.5	96	US-09-540-236-2315	Sequence 2315, App							
172	6	2.5	108	US-09-543-661A-6983	Sequence 6983, App							
173	6	2.5										

## ALIGNMENTS

## RESULT 1

US-09-589-927-6  
; Sequence 6, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 268  
; TYPE: PPT  
; ORGANISM: Human  
US-09-589-927-6

Query Match 66.8%; Score 163; DB 4; Length 268;  
Best Local Similarity 100.0%; Pred. No. 2.7e-159;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	5	KRGDSGSPATWTTGTFVTRHSQTATPSCPECTVPLYSGLFVQGNORAGQDGLTL	64
DB	29	KRGDSGSPATWTTGTFVTRHSQTATPSCPECTVPLYSGLFVQGNORAGQDGLTL	58
QY	65	GSCILQRTFTTMBPFLFCNVNDVNCNFAASNDYSYWLSTPALMPMNPAPITGTRALEPYSRCTV	124
DB	89	GSCILQRTFTTMBPFLFCNVNDVNCNFAASNDYSYWLSTPALMPMNPAPITGTRALEPYSRCTV	148
QY	125	CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFMFTSAGSEG	167
DB	149	CEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPFMFTSAGSEG	191

## RESULT 2

US-09-277-665-6  
; Sequence 6, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:

```
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-6

Query Match
Best Local Similarity 66.8%; Score 163; DB 4; Length 268;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRGDGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 64
Db 29 KRGDGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 88
QY 65 GSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYISRCTV 124
Db 89 GSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYISRCTV 148
QY 125 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 167
Db 149 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 191

RESULT 3
US-09-589-987-6
; Sequence 6, Application US/095898987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-6

Query Match
Best Local Similarity 66.8%; Score 163; DB 4; Length 268;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRGDGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 64
Db 29 KRGDGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQDGLGL 88
QY 65 GSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYISRCTV 124
Db 89 GSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYISRCTV 148
QY 125 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 167
Db 149 CEGPATAIAVHSQTTDIPPCPHGWISLWKGFSPIMFTSAGSEG 191

RESULT 4
US-09-512-563C-46
; Sequence 46, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
```

```
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDV
US-09-512-563C-46

Query Match
Best Local Similarity 65.2%; Score 159; DB 4; Length 211;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQD 60
Db 1 GLKGRGDSGSPATWTRGVFTRHSQTATPSCPECTVPLYSGFSLFVQGNQRAHQD 60
QY 61 LGTLGSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120
Db 61 LGTLGSCLOQRTTTPFLFCNVNDVCFASNDYSYWLSTPALMPMNPATITGRALEPYIS 120
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159
Db 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSPIM 159

RESULT 5
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Query Match
Best Local Similarity 57.8%; Score 141; DB 2; Length 218;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 QTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGLGSCLOQRTTTPFLFCNVNDV 86
Db 1 QTTAIPSCPECTVPLYSGFSLFVQGNQRAHQDGLGLGSCLOQRTTTPFLFCNVNDV 60
QY 87 FASNDYSYWLSTPALMPMNPATITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
Db 61 FASNDYSYWLSTPALMPMNPATITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

RESULT 6
```

```
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Query Match      57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 8.6e-137;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFITMPFLFCNVNDVCN 86
DB 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFITMPFLFCNVNDVCN 60

QY 87 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

QY 147 GWISLWKGFSPIMFTSAGSEG 167
DB 121 GWISLWKGFSPIMFTSAGSEG 141

RESULT 7
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Query Match      57.8%; Score 141; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 8.6e-137;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFITMPFLFCNVNDVCN 86
DB 1 QTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQDLGTLGSCLORFITMPFLFCNVNDVCN 60

QY 87 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 146
DB 61 FASRNDYSYWLSTPALPMNMNAPITGRALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPH 120

QY 147 GWISLWKGFSPIMFTSAGSEG 167
DB 121 GWISLWKGFSPIMFTSAGSEG 141

US-09-512-563C-50
; Sequence 50, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 50
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-IV-V
US-09-512-563C-50

Query Match      25.0%; Score 61; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 4.1e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60

QY 61 L 61
DB 61 L 61

RESULT 8
US-09-512-563C-48
; Sequence 48, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 48
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII
US-09-512-563C-48

Query Match      25.0%; Score 61; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.4e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTTAIPSCPEGTVPVLYSGFSFLVQGNORAHGQD 60

QY 61 L 61
DB 61 L 61

RESULT 9
US-09-512-563C-52
; Sequence 52, Application US/09512563C
```



```
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 52
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPDIII-V
US-09-512-563C-52

Query Match          25.0%; Score 61; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.4e-55;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
DB 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPEGTVPLYSGFSLFVQGNQRAHQD 60
QY 61 L 61
DB 61 L 61

RESULT 11
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; PRIOR FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-08-399-889-24

Query Match          16.0%; Score 39; DB 2; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 12
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
```

```
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; PRIOR FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

Query Match          16.0%; Score 39; DB 3; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 13
US-09-439-897-2
; Sequence 2, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-439-897-2

Query Match          16.0%; Score 39; DB 3; Length 471;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 145
DB 334 MAPITGRALEPYISRCTVCEGPAIAIAVHSQTDTIPPCP 372

RESULT 14
US-09-512-563C-61
; Sequence 61, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 61
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-512-563C-61

Query Match          15.2%; Score 37; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.9e-30;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKGRGDSGSPATWTRGTFVTRHSQTATPSCPEG 37
```

Db 1 GLKRGDGGSPATWTRGFVTRHSQTTPSCPEG 37  
|||||

## RESULT 15

US-09-439-897-65  
; Sequence 65, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 65  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
; OTHER INFORMATION: construct C8 alpha3  
US-09-439-897-65

Query Match 14.8%; Score 36; DB 3; Length 36;

Best Local Similarity 100.0%; Pred. No. 1.1e-29; Indels 0; Gaps 0;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 SLNPERMFKPIPTVKAGELEKIISRCQVCKKQH 244  
|||||

Db 1 SLNPERMFKPIPTVKAGELEKIISRCQVCKKQH 36  
|||||

## RESULT 16

US-09-439-897-63  
; Sequence 63, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 63  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
; OTHER INFORMATION: construct C7 alpha3  
US-09-439-897-63

Query Match 10.7%; Score 26; DB 3; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.5e-19; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKRGDGGSPATWTRGFVTRHS 26  
|||||

Db 1 GLKRGDGGSPATWTRGFVTRHS 26  
|||||

## RESULT 17

US-09-512-563C-26  
; Sequence 26, Application US/09512563C  
; Patent No. 6579969  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; TITLE OF INVENTION: Goodpasture Binding Protein  
; FILE REFERENCE: 98-723-A  
; CURRENT APPLICATION NUMBER: US/09/512,563C

; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/121,483  
; PRIOR FILING DATE: 1999-02-24  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl  
US-09-512-563C-26

Query Match 8.6%; Score 21; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.7e-14; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 KGKRGDGGSPATWTRGFVFT 23  
|||||

Db 1 KGKRGDGGSPATWTRGFVFT 21  
|||||

## RESULT 18

US-09-589-927-2  
; Sequence 2, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-2

Query Match 7.0%; Score 17; DB 4; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.2e-09; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 VCNFASNDYSYWLSTP 100  
|||||

Db 101 VCNFASNDYSYWLSTP 117  
|||||

## RESULT 19

US-09-277-665-2  
; Sequence 2, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-2

Query Match 7.0%; Score 17; DB 4; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.2e-09; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
Db 101 VCNFASRNDYSYWLSTP 117

## RESULT 20

US-09-589-987-2  
; Sequence 2, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-2

Query Match 7.0%; Score 17; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
Db 101 VCNFASRNDYSYWLSTP 117

## RESULT 21

US-09-589-927-10  
; Sequence 10, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-10

Query Match 7.0%; Score 17; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
Db 105 VCNFASRNDYSYWLSTP 121

## RESULT 22

US-09-277-665-10  
; Sequence 10, Application US/09277665  
; Patent No. 6440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-10

Query Match 7.0%; Score 17; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
Db 105 VCNFASRNDYSYWLSTP 121

## RESULT 23

US-09-589-987-10  
; Sequence 10, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; TITLE OF INVENTION: Modify Cell and Tissue Interactions  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-10

Query Match 7.0%; Score 17; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VCNFASRNDYSYWLSTP 100  
Db 105 VCNFASRNDYSYWLSTP 121

## RESULT 24

US-09-439-897-53  
; Sequence 53, Application US/09439897  
; Patent No. 6277556  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C  
; CURRENT APPLICATION NUMBER: US/09/439,897  
; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
; OTHER INFORMATION: construct C2 alpha3  
US-09-439-897-53

Query Match 6.1%; Score 15; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TAIPSCPGTVPPLYS 43  
Db 1 TAIPSCPGTVPPLYS 15

```
RESULT 25
US-09-439-897-57
; Sequence 57, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C4 alpha3
US-09-439-897-57

Query Match
Best Local Similarity 6.1%; Score 15; DB 3; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EKIISRCQVCWKRRH 244
Db 1 EKIISRCQVCWKRRH 15

RESULT 26
US-09-439-897-59
; Sequence 59, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C5 alpha3
US-09-439-897-59

Query Match
Best Local Similarity 6.1%; Score 15; DB 3; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LPMNMAPITGRALE 116
Db 1 LPMNMAPITGRALE 15

RESULT 27
US-09-439-897-61
; Sequence 61, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 61
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C6 alpha3
US-09-439-897-61

Query Match
Best Local Similarity 6.1%; Score 15; DB 3; Length 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 TDIPPCPHGWISLWK 153
Db 1 TDIPPCPHGWISLWK 15

RESULT 28
US-09-512-563C-27
; Sequence 27, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPpepl1a1a9
US-09-512-563C-27

Query Match
Best Local Similarity 5.7%; Score 14; DB 4; Length 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSPATWTTRGPFVT 23
Db 8 GSPATWTTRGPFVT 21

RESULT 29
US-07-621-091G-3
; Sequence 3, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621.091G
```

FILING DATE: 11/30/90  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA: No. 5424408 applicable  
ATTORNEY/AGENT INFORMATION:  
NAME: Barth, Richard S.  
REGISTRATION NUMBER: 28180  
REFERENCE/DOCKET NUMBER: 900983/RB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 972-1400  
TELEFAX: (212) 370-1622  
TELEX: 236268  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acid residues  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: CDNA to mRNA  
DESCRIPTION: Synthetic peptide corresponding to the deduced  
DESCRIPTION: amino acid sequence of the carboxy terminal 12 amino acids of SEQ 1  
FRAGMENT TYPE: C-terminal fragment  
ORIGINAL SOURCE:  
ORGANISM: Human  
INDIVIDUAL ISOLATE: Unknown  
DEVELOPMENTAL STAGE: Unknown  
CELL TYPE: Whole human kidney  
PUBLICATION INFORMATION: No. 5424408e  
US-07-621-091G-3

Query Match 4.9%; Score 12; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 233 ISRCQVCWKRRH 244  
| | | | | | | | | | | | | |  
Db 1 ISRCQVCWKRRH 12  
RESULT 30  
US-08-399-889-3  
Sequence 3, Application US/08399889B  
Patent No. 5973120  
GENERAL INFORMATION:  
APPLICANT: Reeders, Stephen T  
APPLICANT: Morrison, Karen E  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 951263A  
CURRENT APPLICATION NUMBER: US/08/399,889B  
CURRENT FILING DATE: 1995-03-07  
EARLIER APPLICATION NUMBER: 07/621091  
EARLIER FILING DATE: 1990-11-30  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Human  
US-08-399-889-3

Query Match 4.9%; Score 12; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 233 ISRCQVCWKRRH 244  
| | | | | | | | | | | | | |  
Db 1 ISRCQVCWKRRH 12  
RESULT 31  
US-09-167-364-3  
Sequence 3, Application US/09167364  
Patent No. 6007980  
GENERAL INFORMATION:

APPLICANT: Reeders, Stephen T  
APPLICANT: Morrison, Karen E  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 951263B  
CURRENT APPLICATION NUMBER: US/09/167,364  
CURRENT FILING DATE: 1998-10-07  
EARLIER APPLICATION NUMBER: 08/399889  
EARLIER FILING DATE: 1995-03-07  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Human  
US-09-167-364-3

Query Match 4.9%; Score 12; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 233 ISRCQVCWKRRH 244  
| | | | | | | | | | | | | |  
Db 1 ISRCQVCWKRRH 12  
RESULT 32  
US-09-439-897-5  
Sequence 5, Application US/09439897  
Patent No. 6277558  
GENERAL INFORMATION:  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 95-1263-C  
CURRENT APPLICATION NUMBER: US/09/439,897  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 5  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-439-897-5

Query Match 4.9%; Score 12; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 233 ISRCQVCWKRRH 244  
| | | | | | | | | | | | | |  
Db 1 ISRCQVCWKRRH 12  
RESULT 33  
US-09-439-897-55  
Sequence 55, Application US/09439897  
Patent No. 6277558  
GENERAL INFORMATION:  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 95-1263-C  
CURRENT APPLICATION NUMBER: US/09/439,897  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 55  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
OTHER INFORMATION: construct C3 alpha3  
US-09-439-897-55

```
Query Match          4.1%; Score 10; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 SLNPFMRKPI 220
| | | | | | | |
Db 1 SLNPFMRKPI 12

RESULT 34
US-09-589-927-4
; Sequence 4, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-4

Query Match          4.1%; Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQ 137
| | | | | | | |
Db 142 PAIAIAVHSQ 151

RESULT 35
US-09-277-665-4
; Sequence 4, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251-I
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-4

Query Match          4.1%; Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQ 137
| | | | | | | |
Db 142 PAIAIAVHSQ 151

RESULT 36
US-09-589-987-4
; Sequence 4, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
```

```
; TITLE OF INVENTION: Modify Cell and Tissue Interactions
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-4

Query Match          4.1%; Score 10; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIAVHSQ 137
| | | | | | | |
Db 142 PAIAIAVHSQ 151

RESULT 37
US-09-512-563C-44
; Sequence 44, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII-IV-V
; OTHER INFORMATION: derived peptide
US-09-512-563C-44

Query Match          3.3%; Score 8; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 QRAHGQDL 61
| | | | | | | |
Db 1 QRAHGQDL 8

RESULT 38
US-09-512-563C-43
; Sequence 43, Application US/09512563C
; Patent No. 6579969
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Goodpasture Binding Protein
; FILE REFERENCE: 98-723-A
; CURRENT APPLICATION NUMBER: US/09/512,563C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/121,483
; PRIOR FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GPIII derived
```

```
; OTHER INFORMATION: peptide
US-09-512-563C-43
Query Match      3.3%; Score 8; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 QRAHQDGL 61
   |||||
Db 1 QRAHQDGL 8

RESULT 39
US-09-589-927-12
; Sequence 12, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,927
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-927-12
Query Match      3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
   |||||
Db 251 SRCQVCWK 258

RESULT 40
US-09-277-665-12
; Sequence 12, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 94525-1
; CURRENT APPLICATION NUMBER: US/09/277,665
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-277-665-12
Query Match      3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
   |||||
Db 251 SRCQVCWK 258

RESULT 41
US-09-589-987-12
; Sequence 12, Application US/09589987
; Patent No. 6498140
; GENERAL INFORMATION:
; APPLICANT: University of Kansas Medical Center
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to
; FILE REFERENCE: 945251
; CURRENT APPLICATION NUMBER: US/09/589,987
; CURRENT FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Human
US-09-589-987-12
Query Match      3.3%; Score 8; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 241
   |||||
Db 251 SRCQVCWK 258

RESULT 42
US-09-439-897-56
; Sequence 56, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C4 alpha1
US-09-439-897-56
Query Match      2.9%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCWK 240
   |||||
Db 5 SRCQVCWK 11

RESULT 43
US-09-439-897-64
; Sequence 64, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C8 alpha1
US-09-439-897-64
```



```
Query Match      2.9%; Score 7; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 SRCQVCM 240
    |||||
Db 26 SRCQVCM 32

RESULT 44
US-09-134-001C-3093
; Sequence 3093, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 674
; SEQ ID NO 3093
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3093

Query Match      2.9%; Score 7; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSPA 13
    |||||
Db 40 GDSGSPA 46

RESULT 45
US-09-395-689-6
; Sequence 6, Application US/09395689
; Patent No. 6387684
; GENERAL INFORMATION:
; APPLICANT: Hwang, Jaulang
; APPLICANT: Hui, cho-Pat
; APPLICANT: Chen, Tzong-Yueh
; TITLE OF INVENTION: TOPOLISOMERASE 1-MEDIATED DNA DELIVERY
; FILE REFERENCE: 089191/024001
; CURRENT APPLICATION NUMBER: US/09/395,689
; CURRENT FILING DATE: 1999-09-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-395-689-6

Query Match      2.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
    |||||
Db 88 ALASPGS 94

RESULT 46
US-09-411-578-1
; Sequence 1, Application US/09411578
```

```
; Patent No. 6203801
; GENERAL INFORMATION:
; APPLICANT: Schaap, Theodorus C
; APPLICANT: Kuiper, Catharina M
; APPLICANT: Vermeulen, Arnoldus N
; TITLE OF INVENTION: Coccidiosis Vaccines
; FILE REFERENCE: schaap
; CURRENT APPLICATION NUMBER: US/09/411,578
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 98203384.7
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 98203457.1
; EARLIER FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Eimeria tenella
US-09-411-578-1

Query Match      2.9%; Score 7; DB 3; Length 214;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
    |||||
Db 14 ALEPYIS 20

RESULT 47
US-09-749-233-1
; Sequence 1, Application US/09749233
; Patent No. 6680061
; GENERAL INFORMATION:
; APPLICANT: Schaap, Theodorus C
; APPLICANT: Kuiper, Catharina M
; APPLICANT: Vermeulen, Arnoldus N
; TITLE OF INVENTION: Coccidiosis Vaccines
; FILE REFERENCE: schaap
; CURRENT APPLICATION NUMBER: US/09/749,233
; CURRENT FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/411,578
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: 98203457.1
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Eimeria tenella
US-09-749-233-1

Query Match      2.9%; Score 7; DB 4; Length 214;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120
    |||||
Db 14 ALEPYIS 20

RESULT 48
US-08-486-099-96
; Sequence 96, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
```

```

; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; TITLE OF INVENTION: B VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-486-099-96

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 49
US-08-360-107A-106
; Sequence 106, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-360-107A-106

Query Match 2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171
Db 29 SEGTGQA 35

RESULT 50
US-08-484-223B-96
; Sequence 96, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE

```

INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-223B-96

Query Match 2.9%; Score 7; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 51

US-08-919-597-96  
Sequence 96, Application US/08919597  
Patent No. 6054265  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/919,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/470,896  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-919-597-96

Query Match 2.9%; Score 7; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 52

US-08-475-668A-96  
Sequence 96, Application US/08475668A  
Patent No. 6060065  
GENERAL INFORMATION:  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10038-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,668A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-475-668A-96

Query Match 2.9%; Score 7; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 53

US-08-485-551A-96  
Sequence 96, Application US/08485551A  
Patent No. 6068973  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE

```

; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,551A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-551A-96

Query Match          2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      165 SEGTGQA 171
        |||||
Db       29 SEGTGQA 35

RESULT 54
US-08-471-913A-96
; Sequence 96, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30

Query Match          2.9%; Score 7; DB 3; Length 221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      165 SEGTGQA 171
        |||||
Db       29 SEGTGQA 35

RESULT 55
US-08-485-264A-96
; Sequence 96, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids

```

TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-485-264A-96

Query Match 2.9%; Score 7; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 56

US-08-474-349A-96  
Sequence 96, Application US/08474349A  
Patent No. 633395

## GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway Jr., Stephen R.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA  
TITLE OF INVENTION: VIRUS TRANSMISSION  
NUMBER OF SEQUENCES: 517  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,349A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-024  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-474-349A-96

Query Match 2.9%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 57

US-08-255-208A-32  
Sequence 32, Application US/08255208A  
Patent No. 644656

## GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway Jr., Stephen R.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/255,208A  
FILING DATE: 07-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-255-208A-32

Query Match 2.9%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 58

US-08-470-896-96  
Sequence 96, Application US/08470896  
Patent No. 647955

## GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/470,896  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-470-896-96

Query Match 2.9%, Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 29 SEGTGQA 35

RESULT 59  
US-08-485-546A-96  
Sequence 96, Application US/08485546A  
Patent No. 6518013  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE  
FUSION  
TITLE OF INVENTION: FUSION-ASSOCIATED TRANSMISSION  
NUMBER OF SEQUENCES: 214  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,546A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-485-546A-96

Query Match 2.9%, Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 29 SEGTGQA 35

RESULT 60  
PCT-US94-01149-2  
Sequence 2, Application PC/TUS9401149  
GENERAL INFORMATION:  
APPLICANT: Shatzman, Allan  
APPLICANT: Scott, Miller  
APPLICANT: Dillon, Susan B.  
APPLICANT: Kane, James  
TITLE OF INVENTION: Vaccinal Polypeptides  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate  
STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/01149  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 149,150  
FILING DATE: 05-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 013,415  
FILING DATE: 01-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 108,914  
FILING DATE: 18-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 837,773  
FILING DATE: 18-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 751,896  
FILING DATE: 30-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 387,200  
FILING DATE: 28-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 238,801  
FILING DATE: 02-NOV-1988

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 645,732  
;; FILING DATE: 30-AUG-1984  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Baumeister, Kirk  
;; REGISTRATION NUMBER: 33,833  
;; REFERENCE/DOCKET NUMBER: P50134 PCT  
;; TELEPHONE: 215-270-5096  
;; TELEFAX: 215-270-5090  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 221 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US94-01149-2

Query Match 2.9%; Score 7; DB 5; Length 221;  
Best Local Similarity 100.0%; Pred.No.35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
|||  
Db 29 SEGTGQA 35

RESULT 61  
PCT-US94-01149-4  
;; Sequence 4, Application PC/TUS9401149  
;; GENERAL INFORMATION:  
;; APPLICANT: Shatzman, Allan  
;; APPLICANT: Scott, Miller  
;; APPLICANT: Dillon, Susan B.  
;; APPLICANT: Kane, James  
;; TITLE OF INVENTION: Vaccinal Polypeptides  
;; NUMBER OF SEQUENCES: 72  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate  
;; ADDRESSEE: Patents  
;; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
;; CITY: King of Prussia  
;; STATE: Pennsylvania  
;; COUNTRY: USA  
;; ZIP: 19406-2799  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US94/01149  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 149,150  
;; FILING DATE: 05-NOV-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 013,415  
;; FILING DATE: 01-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 108,914  
;; FILING DATE: 18-AUG-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 837,773  
;; FILING DATE: 18-FEB-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 751,896  
;; FILING DATE: 30-AUG-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 387,200  
;; FILING DATE: 28-JUL-1989  
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 238,801  
;; FILING DATE: 02-NOV-1988  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 645,732  
;; FILING DATE: 30-AUG-1984  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Baumeister, Kirk  
;; REGISTRATION NUMBER: 33,833  
;; REFERENCE/DOCKET NUMBER: P50134 PCT  
;; TELEPHONE: 215-270-5096  
;; TELEFAX: 215-270-5090  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 221 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US94-01149-4

Query Match 2.9%; Score 7; DB 5; Length 221;  
Best Local Similarity 100.0%; Pred.No.35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
|||  
Db 29 SEGTGQA 35

RESULT 62  
PCT-US94-01149-59  
;; Sequence 59, Application PC/TUS9401149  
;; GENERAL INFORMATION:  
;; APPLICANT: Shatzman, Allan  
;; APPLICANT: Scott, Miller  
;; APPLICANT: Dillon, Susan B.  
;; APPLICANT: Kane, James  
;; TITLE OF INVENTION: Vaccinal Polypeptides  
;; NUMBER OF SEQUENCES: 72  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate  
;; ADDRESSEE: Patents  
;; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
;; CITY: King of Prussia  
;; STATE: Pennsylvania  
;; COUNTRY: USA  
;; ZIP: 19406-2799  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US94/01149  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 149,150  
;; FILING DATE: 05-NOV-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 013,415  
;; FILING DATE: 01-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 108,914  
;; FILING DATE: 18-AUG-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 837,773  
;; FILING DATE: 18-FEB-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 751,896  
;; FILING DATE: 30-AUG-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 387,200



;  
; FILING DATE: 28-JUL-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 238,801  
; FILING DATE: 02-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 238,801  
; FILING DATE: 02-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 645,732  
; FILING DATE: 30-AUG-1984  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833  
; REFERENCE/DOCKET NUMBER: P50134 PCT  
; TELEPHONE: 215-270-5096  
; TELEFAX: 215-270-5090  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US94-01149-59

Query Match 2.9%; Score 7; DB 5; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 63

PCT-US94-01149-60  
; Sequence 60, Application PC/TUS9401149  
; GENERAL INFORMATION:  
; APPLICANT: Shatzman, Allan  
; APPLICANT: Scott, Miller  
; APPLICANT: Dillon, Susan B.  
; APPLICANT: Kane, James  
; TITLE OF INVENTION: Vaccinal Polypeptides  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation - Corporate  
; ADDRESSEE: Patents  
; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/01149  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 149,150  
; FILING DATE: 05-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 013,415  
; FILING DATE: 01-FEB-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 108,914  
; FILING DATE: 18-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 837,773  
; FILING DATE: 18-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 751,896  
; FILING DATE: 30-AUG-1991

;  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 387,200  
; FILING DATE: 28-JUL-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 238,801  
; FILING DATE: 02-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 645,732  
; FILING DATE: 30-AUG-1984  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833  
; REFERENCE/DOCKET NUMBER: P50134 PCT  
; TELEPHONE: 215-270-5096  
; TELEFAX: 215-270-5090  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US94-01149-60

Query Match 2.9%; Score 7; DB 5; Length 221;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
Db 29 SEGTGQA 35

## RESULT 64

PCT-US94-01149-8  
; Sequence 8, Application PC/TUS9401149  
; GENERAL INFORMATION:  
; APPLICANT: Shatzman, Allan  
; APPLICANT: Scott, Miller  
; APPLICANT: Dillon, Susan B.  
; APPLICANT: Kane, James  
; TITLE OF INVENTION: Vaccinal Polypeptides  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation - Corporate  
; ADDRESSEE: Patents  
; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/01149  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 149,150  
; FILING DATE: 05-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 013,415  
; FILING DATE: 01-FEB-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 108,914  
; FILING DATE: 18-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 837,773  
; FILING DATE: 18-FEB-1992  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 751,896  
FILING DATE: 30-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 387,200  
FILING DATE: 28-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 238,801  
FILING DATE: 02-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 645,732  
FILING DATE: 30-AUG-1984  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50134 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-270-5096  
TELEFAX: 215-270-5090  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
PCT-US94-01149-8

Query Match 2.9%; Score 7; DB 5; Length 222;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
DB 29 SEGTGQA 35

RESULT 65  
US-08-928-692-31  
Sequence 31, Application US/08928692  
Patent No. 5958727  
GENERAL INFORMATION:  
APPLICANT: Brody, Howard  
APPLICANT: Yaver, Deborah S.  
APPLICANT: Lamsa, Michael  
APPLICANT: Hansen, Kim  
TITLE OF INVENTION: Methods for Modifying the Production of  
TITLE OF INVENTION: a Polypeptide  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,692  
FILING DATE: 12-SEPT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4944.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
PCT-US94-01149-8

LENGTH: 233 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5958727e  
US-08-928-692-31

Query Match 2.9%; Score 7; DB 2; Length 233;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120  
DB 39 ALEPYIS 45

RESULT 66  
US-09-339-972-31  
Sequence 31, Application US/09339972  
Patent No. 6323002  
GENERAL INFORMATION:  
APPLICANT: Brody, Howard  
APPLICANT: Yaver, Deborah S.  
APPLICANT: Lamsa, Michael  
APPLICANT: Hansen, Kim  
TITLE OF INVENTION: Methods for Modifying the Production of  
TITLE OF INVENTION: a Polypeptide  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6323002o No. 6323002disk of No. 6323002th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/339,972  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/928,692  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4944.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 233 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6323002e  
US-09-339-972-31

Query Match 2.9%; Score 7; DB 4; Length 233;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ALEPYIS 120  
DB 39 ALEPYIS 45

RESULT 67

5223425-2  
; Patent No. 5223425  
; APPLICANT: FLIER, JEFFREY S.; SPIEGELMAN, BRUCE M.; ROSEN,  
; BARRY M.; WHITE, TYLER R.  
; TITLE OF INVENTION: DNA ENCODING HUMAN ADIPSIN WITH COMPLEMENT  
; D ACTIVITY  
; NUMBER OF SEQUENCES: 19  
; CURRENT APPLICATION NUMBER: US/07/277,963  
; FILING DATE: 30-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 34,203  
; FILING DATE: 02-APR-1987  
; SEQ ID NO:2:  
; LENGTH: 259  
5223425-2

Query Match 2.9%; Score 7; DB 6; Length 259;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSGP 12  
|||||  
Db 206 RGDGSGP 212

RESULT 68  
US-09-589-927-8  
; Sequence 8, Application US/09589927  
; Patent No. 6432706  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,927  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-927-8

Query Match 2.9%; Score 7; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180  
|||||  
Db 188 SPGSCLE 194

RESULT 69  
US-09-277-665-8  
; Sequence 8, Application US/09277665  
; Patent No. 8440729  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 94525-1  
; CURRENT APPLICATION NUMBER: US/09/277,665  
; CURRENT FILING DATE: 1999-03-26  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-277-665-8

Query Match 2.9%; Score 7; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180  
|||||  
Db 188 SPGSCLE 194

RESULT 70  
US-09-589-987-8  
; Sequence 8, Application US/09589987  
; Patent No. 6498140  
; GENERAL INFORMATION:  
; APPLICANT: University of Kansas Medical Center  
; TITLE OF INVENTION: The Use of Isolated Domains of Type IV Collagen to  
; FILE REFERENCE: 945251  
; CURRENT APPLICATION NUMBER: US/09/589,987  
; CURRENT FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Human  
US-09-589-987-8

Query Match 2.9%; Score 7; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 SPGSCLE 180  
|||||  
Db 188 SPGSCLE 194

RESULT 71  
5223425-10  
; Patent No. 5223425  
; APPLICANT: FLIER, JEFFREY S.; SPIEGELMAN, BRUCE M.; ROSEN,  
; BARRY M.; WHITE, TYLER R.  
; TITLE OF INVENTION: DNA ENCODING HUMAN ADIPSIN WITH COMPLEMENT  
; D ACTIVITY  
; NUMBER OF SEQUENCES: 19  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/277,963  
; FILING DATE: 30-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 34,203  
; FILING DATE: 02-APR-1987  
; SEQ ID NO:10:  
; LENGTH: 260  
5223425-10

Query Match 2.9%; Score 7; DB 6; Length 260;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSGP 12  
|||||  
Db 207 RGDGSGP 213

RESULT 72  
PCT-US94-01149-10  
; Sequence 10, Application PC/TUS9401149  
; GENERAL INFORMATION:  
; APPLICANT: Shatzman, Allan  
; APPLICANT: Scott, Miller  
; APPLICANT: Dillon, Susan B.  
; APPLICANT: Kane, James  
; TITLE OF INVENTION: Vaccinal Polypeptides

;; NUMBER OF SEQUENCES: 72  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SmithKline Beecham Corporation - Corporate  
;; ADDRESSEE: Patents  
;; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road  
;; CITY: King of Prussia  
;; STATE: Pennsylvania  
;; COUNTRY: USA  
;; ZIP: 19406-2799  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US94/01149  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 149,150  
;; FILING DATE: 05-NOV-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 013,415  
;; FILING DATE: 01-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 108,914  
;; FILING DATE: 18-AUG-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 837,773  
;; FILING DATE: 18-FEB-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 751,896  
;; FILING DATE: 30-AUG-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 387,200  
;; FILING DATE: 26-JUL-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 238,801  
;; FILING DATE: 02-NOV-1988  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 645,732  
;; FILING DATE: 30-AUG-1984  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Baumeister, Kirk  
;; REGISTRATION NUMBER: 33,833  
;; REFERENCE/DOCKET NUMBER: P50134 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-270-5096  
;; TELEFAX: 215-270-5090  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 306 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; PCT-US94-01149-10

Query Match 2.9%; Score 7; DB 5; Length 306;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 SEG7GQA 171  
Db 114 SEG7GQA 120

RESULT 73  
US-08-840-713-4  
; Sequence 4, Application US/08840713  
; Patent No. 6498233  
; GENERAL INFORMATION:  
; APPLICANT: WELLS, Winfried, Dr.  
; ATTORNEY/AGENT INFORMATION:  
; FILING DATE: April 29, 1993  
; APPLICATION NUMBER: US/08/229,781  
; FILING DATE: April 19, 1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/054,016  
; FILING DATE: April 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:

;; TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM  
;; NUMBER OF SEQUENCES: 58  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
;; STREET: 655 15th St., N.W., Suite 330 - G St. Lobby  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-5701  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/840,713  
;; FILING DATE: 25-APR-1997  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kitts, Monica Chin  
;; REGISTRATION NUMBER: 36,105  
;; REFERENCE/DOCKET NUMBER: 1614-7014  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 638-5000  
;; TELEFAX: (202) 638-4810  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 342 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-840-713-4

Query Match 2.9%; Score 7; DB 4; Length 342;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 171 ALASPGS 177  
Db 136 ALASPGS 142

RESULT 74  
US-08-229-781-58  
; Sequence 58, Application US/08229781  
; Patent No. 5589174  
; GENERAL INFORMATION:  
; APPLICANT: Yoshinobu OKUNO et al.  
; TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/229,781  
; FILING DATE: April 19, 1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/054,016  
; FILING DATE: April 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:

## INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:  
LENGTH: 347 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:

ORGANISM:  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:

## LIBRARY:

CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:

## FEATURE:

NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:

## AUTHORS:

TITLE:

JOURNAL:

VOLUME:

ISSUE:

PAGES:

DATE:

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-229-781-58

Query Match 2.9%; Score 7; DB 1; Length 347;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 155 SEGTGQA 161

## RESULT 75

US-08-630-918-58  
Sequence 58, Application US/08630918  
Patent No. 5631350  
GENERAL INFORMATION:

APPLICANT: Yoshinobu OKUNO et al.

TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,918  
FILING DATE: April 5, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/229,781  
FILING DATE: April 19, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/054,016  
FILING DATE: April 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:

## INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:  
LENGTH: 347 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-630-918-58

Query Match 2.9%; Score 7; DB 1; Length 347;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 155 SEGTGQA 161

## RESULT 76

US-09-004-422-58  
Sequence 58, Application US/09004422  
Patent No. 6337070  
GENERAL INFORMATION:

APPLICANT: Yoshinobu OKUNO et al.

TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING

ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/004,422

FILING DATE: January 8, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/443,862

FILING DATE: May 22, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,781

FILING DATE: April 19, 1994

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/054,016  
;; FILING DATE: April 29, 1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Warren M. Cheek, Jr.  
;; REGISTRATION NUMBER: 33,367  
;; REFERENCE/DOCKET NUMBER:  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-721-8200  
;; TELEFAX: 202-721-8250  
;;  
;; INFORMATION FOR SEQ ID NO: 58:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 347 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL:  
;; ANTI-SENSE:  
;; FRAGMENT TYPE:  
;; ORIGINAL SOURCE:  
;; ORGANISM:  
;; STRAIN:  
;; INDIVIDUAL ISOLATE:  
;; DEVELOPMENTAL STAGE:  
;; HAPLOTYPE:  
;; TISSUE TYPE:  
;; CELL TYPE:  
;; CELL LINE:  
;; ORGANELLE:  
;; IMMEDIATE SOURCE:  
;; LIBRARY:  
;; CLONE:  
;; POSITION IN GENOME:  
;; CHROMOSOME/SEGMENT:  
;; MAP POSITION:  
;; UNITS:  
;; FEATURE:  
;; NAME/KEY:  
;; LOCATION:  
;; IDENTIFICATION METHOD:  
;; OTHER INFORMATION:  
;; PUBLICATION INFORMATION:  
;; AUTHORS:  
;; TITLE:  
;; JOURNAL:  
;; VOLUME:  
;; ISSUE:  
;; PAGES:  
;; DATE:  
;; DOCUMENT NUMBER:  
;; FILING DATE:  
;; PUBLICATION DATE:  
;; RELEVANT RESIDUES IN SEQ ID NO:  
US-09-004-422-58

Query Match 2.9%; Score 7; DB 4; Length 347;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 155 SEGTGQA 161

RESULT 77  
US-09-973-963-4  
; Sequence 4, Application US/09973963  
; Patent No. 6653102  
; GENERAL INFORMATION:  
; APPLICANT: Roch, Jean-Marc  
; APPLICANT: Bartel, Paul L.

;; APPLICANT: Heichman, Karen  
;; TITLE OF INVENTION: Protein-Protein Interactions in Neurodegenerative  
;; TITLE OF INVENTION: Diseases  
;; FILE REFERENCE: Protein Interactions in ND  
;; CURRENT APPLICATION NUMBER: US/09/973,963  
;; CURRENT FILING DATE: 2001-10-11  
;; PRIOR APPLICATION NUMBER: US 60/240,790  
;; PRIOR FILING DATE: 2000-10-17  
;; PRIOR APPLICATION NUMBER: US 60/304,775  
;; PRIOR FILING DATE: 2001-07-13  
;; NUMBER OF SEQ ID NOS: 8  
;; SOFTWARE: Patent in Ver. 2.0  
;; SEQ ID NO 4  
;; LENGTH: 372  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-973-963-4  
  
Query Match 2.9%; Score 7; DB 4; Length 372;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 9 SGSPATW 15  
Db 55 SGSPATW 61  
  
RESULT 78  
US-09-046-992-4  
; Sequence 4, Application US/09046992  
; Patent No. 6140066  
; GENERAL INFORMATION:  
; APPLICANT: Lorberbaum-Galski, Haya  
; APPLICANT: Yarkoni, Shai  
; APPLICANT: Ben-Yehudah, Ahmi  
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS  
; TITLE OF INVENTION: USING A CHIMERIC TOXIN  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/046,992  
; FILING DATE: 24-MAR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Poissant, Brian M  
; REGISTRATION NUMBER: 28,462  
; REFERENCE/DOCKET NUMBER: 9457-0013-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal

US-09-046-992-4

Query Match 2.9%; Score 7; DB 3; Length 396;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
| | | | |  
Db 98 ALASPGS 104

RESULT 79

US-08-391-259-2

; Sequence 2, Application US/08391259  
; Patent No. 5621078

GENERAL INFORMATION:

APPLICANT: Riemen, Mark W

APPLICANT: Stirdivant, Steven M

TITLE OF INVENTION: Modified PE40

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck &amp; Co., Inc.

STREET: 126 Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: U.S.

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/391,259

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/120,698

FILING DATE:

APPLICATION NUMBER: US/07/879,037

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Grassler, Frank P

REGISTRATION NUMBER: 31,164

REFERENCE/DOCKET NUMBER: 178791A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 594-3462

TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 420 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-391-259-2

Query Match

2.9%; Score 7; DB 1; Length 420;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
| | | | |  
Db 122 ALASPGS 128

RESULT 80

US-08-391-259-7

; Sequence 7, Application US/08391259  
; Patent No. 5621078

GENERAL INFORMATION:

APPLICANT: Riemen, Mark W

APPLICANT: Stirdivant, Steven M

; TITLE OF INVENTION: Modified PE40  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: 126 Lincoln Avenue  
; CITY: Rahway  
; STATE: New Jersey  
; COUNTRY: U.S.  
; ZIP: 07065  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/391,259  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,698  
; FILING DATE:  
; APPLICATION NUMBER: US/07/879,037  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Grassler, Frank P  
; REGISTRATION NUMBER: 31,164  
; REFERENCE/DOCKET NUMBER: 178791A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-3462  
; TELEFAX: (908) 594-4720  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 420 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-391-259-7

Query Match

2.9%; Score 7; DB 1; Length 420;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
| | | | |  
Db 122 ALASPGS 128

RESULT 81

US-08-391-259-10

; Sequence 10, Application US/08391259  
; Patent No. 5621078

GENERAL INFORMATION:

APPLICANT: Riemen, Mark W

APPLICANT: Stirdivant, Steven M

TITLE OF INVENTION: Modified PE40

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck &amp; Co., Inc.

STREET: 126 Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: U.S.

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/391,259

FILING DATE:

CLASSIFICATION: 530



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,698
; FILING DATE:
; APPLICATION NUMBER: US/07/879,037
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-391-259-10

Query Match 2.9%; Score 7; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 122 ALASPGS 128

RESULT 82
US-08-391-259-11
; Sequence 11, Application US/08391259
; Patent No. 5621078
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/391,259
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,698
; FILING DATE:
; APPLICATION NUMBER: US/07/879,037
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
US-08-391-259-11

Query Match 2.9%; Score 7; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 122 ALASPGS 128

RESULT 83
US-08-839-425-2
; Sequence 2, Application US/08839425
; Patent No. 5912322
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Power Macintosh 6.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,425
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Grassler, Frank P
; REGISTRATION NUMBER: 31,164
; REFERENCE/DOCKET NUMBER: 178791A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3462
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-839-425-2

Query Match 2.9%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 122 ALASPGS 128

RESULT 84
US-08-839-425-7
; Sequence 7, Application US/08839425
; Patent No. 5912322
; GENERAL INFORMATION:
; APPLICANT: Riemen, Mark W
; APPLICANT: Stirdivant, Steven M
; TITLE OF INVENTION: Modified PE40
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.

```

STREET: 126 Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: U.S.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Power Macintosh 6.0.1  
SOFTWARE: Microsoftword 6.0.1  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Grassler, Frank P  
REGISTRATION NUMBER: 31,164  
REFERENCE/DOCKET NUMBER: 178791A  
TELEPHONE: (908) 594-3462  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 420 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-839-425-7

Query Match 2.9%; Score 7; DB 2; Length 420;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
DB 122 ALASPGS 128

RESULT 85  
US-08-839-425-10  
Sequence 10, Application US/08839425  
Patent No. 5912322  
GENERAL INFORMATION:  
APPLICANT: Riemen, Mark W  
APPLICANT: Stirdivant, Steven M  
TITLE OF INVENTION: Modified PE40  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: 126 Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: U.S.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Power Macintosh 6.0.1  
SOFTWARE: Microsoftword 6.0.1  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Grassler, Frank P  
REGISTRATION NUMBER: 31,164  
REFERENCE/DOCKET NUMBER: 178791A  
TELEPHONE: (908) 594-3462  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 420 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-839-425-10

Query Match 2.9%; Score 7; DB 2; Length 420;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
DB 122 ALASPGS 128

RESULT 86  
US-08-839-425-11  
Sequence 11, Application US/08839425  
Patent No. 5912322  
GENERAL INFORMATION:  
APPLICANT: Riemen, Mark W  
APPLICANT: Stirdivant, Steven M  
TITLE OF INVENTION: Modified PE40  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: 126 Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: U.S.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Power Macintosh 6.0.1  
SOFTWARE: Microsoftword 6.0.1  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Grassler, Frank P  
REGISTRATION NUMBER: 31,164  
REFERENCE/DOCKET NUMBER: 178791A  
TELEPHONE: (908) 594-3462  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 420 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-839-425-11

Query Match 2.9%; Score 7; DB 2; Length 420;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
DB 122 ALASPGS 128

RESULT 87  
US-08-840-713-6  
Sequence 6, Application US/08840713  
Patent No. 6498233  
GENERAL INFORMATION:  
APPLICANT: WELLS, Winfried, Dr.  
APPLICANT: FOYMINAYA, Jesus

;; TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM  
;; NUMBER OF SEQUENCES: 58  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
;; STREET: 655 15th St., N.W., Suite 330 - G St. Lobby  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-5701  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/840,713  
;; FILING DATE: 25-APR-1997  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kitts, Monica Chin  
;; REGISTRATION NUMBER: 36,105  
;; REFERENCE/DOCKET NUMBER: 1614-7014  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 638 - 5000  
;; TELEFAX: (202) 638 - 4810  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 421 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-840-713-6

Query Match 2.9%; Score 7; DB 4; Length 421;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 215 ALASPGS 221

RESULT 88  
US-08-484-438-42  
;; Sequence 42, Application US/08484438  
;; Patent No. 5811098  
;; Patent No. 5811098 5780031  
;; GENERAL INFORMATION:  
;; APPLICANT: Plowman, Gregory D.  
;; APPLICANT: Culouscou, Jean-Michel  
;; APPLICANT: Shoyab, Mohammed  
;; APPLICANT: Siegall, Clay B.  
;; APPLICANT: Heilstr m, Ingegerd  
;; APPLICANT: Heilstr m, Karl E.  
;; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE  
;; NUMBER OF SEQUENCES: 42  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/484,438  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/323,442  
;; FILING DATE: 14-OCT-1994  
;; APPLICATION NUMBER: US 08/150,704  
;; FILING DATE: 10-NOV-1993  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/981,165  
;; FILING DATE: 24-NOV-1992  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Misrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 5624-230  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 42:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 462 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-484-438-42

Query Match 2.9%; Score 7; DB 2; Length 462;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 164 ALASPGS 170

RESULT 89  
US-08-840-713-2  
;; Sequence 2, Application US/08840713  
;; Patent No. 6498233  
;; GENERAL INFORMATION:  
;; APPLICANT: WELS, Winfried, Dr.  
;; APPLICANT: FOYMINAYA, Jesus  
;; TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM  
;; NUMBER OF SEQUENCES: 58  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
;; STREET: 655 15th St., N.W., Suite 330 - G St. Lobby  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-5701  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/840,713  
;; FILING DATE: 25-APR-1997  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kitts, Monica Chin  
;; REGISTRATION NUMBER: 36,105  
;; REFERENCE/DOCKET NUMBER: 1614-7014  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 638 - 5000  
;; TELEFAX: (202) 638 - 4810  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 530 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-840-713-2

Query Match 2.9%; Score 7; DB 4; Length 530;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
|||||||  
DB 324 ALASPGS 330

RESULT 90  
US-09-506-286B-8  
; Sequence 8, Application US/09506286B  
; Patent No. 6482414  
; GENERAL INFORMATION:  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; APPLICANT: The University of Pittsburgh, of the Commonwealth  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C2  
; CURRENT APPLICATION NUMBER: US/09/506,286B  
; CURRENT FILING DATE: 2000-02-16  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1999-08-12  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-09-506-286B-8

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||||  
DB 373 SEGTGQA 379

RESULT 91  
US-09-506-286B-11  
; Sequence 11, Application US/09506286B  
; Patent No. 6482414  
; GENERAL INFORMATION:  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; APPLICANT: The University of Pittsburgh, of the Commonwealth  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C2  
; CURRENT APPLICATION NUMBER: US/09/506,286B  
; CURRENT FILING DATE: 2000-02-16  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1999-08-12  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-09-506-286B-11

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||||

DB 373 SEGTGQA 379

RESULT 92  
US-09-762-861B-8  
; Sequence 8, Application US/09762861B  
; Patent No. 6579528  
; GENERAL INFORMATION:  
; APPLICANT: The University of Pittsburgh - of the Commonwealth System of Higher  
; APPLICANT: Education  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C1-PUS (formerly HKZ-033CPUS)  
; CURRENT APPLICATION NUMBER: US/09/762,861B  
; CURRENT FILING DATE: 2001-02-13  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: PCT/US99/18583  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-09-762-861B-8

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||||  
DB 373 SEGTGQA 379

RESULT 93  
US-09-762-861B-11  
; Sequence 11, Application US/09762861B  
; Patent No. 6579528  
; GENERAL INFORMATION:  
; APPLICANT: The University of Pittsburgh - of the Commonwealth System of Higher  
; APPLICANT: Education  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C1-PUS (formerly HKZ-033CPUS)  
; CURRENT APPLICATION NUMBER: US/09/762,861B  
; CURRENT FILING DATE: 2001-02-13  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: PCT/US99/18583  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-09-762-861B-11

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 SEGTGQA 171  
|||||||  
DB 373 SEGTGQA 379

RESULT 94  
US-10-065-133A-8

; Sequence 8, Application US/10065133A  
; Patent No. 6685946  
; GENERAL INFORMATION:  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C2-1  
; CURRENT APPLICATION NUMBER: US/10/065,133A  
; CURRENT FILING DATE: 2002-12-10  
; PRIOR APPLICATION NUMBER: PCT/US99/18583  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-10-065-133A-8

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 373 SEGTGQA 379

RESULT 95  
US-10-065-133A-11  
; Sequence 11, Application US/10065133A  
; Patent No. 6685946  
; GENERAL INFORMATION:  
; APPLICANT: Dowling, Patricia W.  
; APPLICANT: Youngner, Julius S.  
; TITLE OF INVENTION: COLD-ADAPTED EQUINE INFLUENZA VIRUSES  
; FILE REFERENCE: EQ-1-C2-1  
; CURRENT APPLICATION NUMBER: US/10/065,133A  
; CURRENT FILING DATE: 2002-12-10  
; PRIOR APPLICATION NUMBER: PCT/US99/18583  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 09/133,921  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 565  
; TYPE: PRT  
; ORGANISM: Equine influenza virus H3N8  
US-10-065-133A-11

Query Match 2.9%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 373 SEGTGQA 379

RESULT 96  
US-09-232-468A-22  
; Sequence 22, Application US/09232468A  
; Patent No. 6207165  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES  
; FILE REFERENCE: 454313-2230  
; CURRENT APPLICATION NUMBER: US/09/232,468A  
; CURRENT FILING DATE: 1999-01-05

; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 566  
; TYPE: PRT  
; ORGANISM: swine influenza virus  
US-09-232-468A-22

Query Match 2.9%; Score 7; DB 3; Length 566;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 374 SEGTGQA 380

RESULT 97  
US-09-784-984B-53  
; Sequence 53, Application US/09784984B  
; Patent No. 6576243  
; GENERAL INFORMATION:  
; APPLICANT: Meriel Ltd.  
; APPLICANT: Audonnet, Jean-Christophe  
; APPLICANT: Bouchardon, Annabelle  
; APPLICANT: Baudu, Philippe  
; APPLICANT: Riviere, Michael  
; TITLE OF INVENTION: Polynucleotide Vaccine Formula Against Porcine Reproductive and  
; TITLE OF INVENTION: Respiratory Pathologies  
; FILE REFERENCE: 454313-2230.1  
; CURRENT APPLICATION NUMBER: US/09/784,984B  
; CURRENT FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: FR 96/09338  
; PRIOR FILING DATE: 1996-07-19  
; PRIOR APPLICATION NUMBER: PCT/FR97/01313  
; PRIOR FILING DATE: 1997-07-15  
; PRIOR APPLICATION NUMBER: US 6,207,165  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 53  
; LENGTH: 566  
; TYPE: PRT  
; ORGANISM: Swine Influenza Virus  
US-09-784-984B-53

Query Match 2.9%; Score 7; DB 4; Length 566;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 374 SEGTGQA 380

RESULT 98  
US-08-453-848-7  
; Sequence 7, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta

```
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/453,848
  FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/120,607
  FILING DATE: 13-SEPT-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Pabst, Patrea L.
  REGISTRATION NUMBER: 31,284
  REFERENCE/DOCKET NUMBER: MGS101CIP
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (404)-873-8794
    TELEFAX: (404)-873-8795
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 570 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: Peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE: N-terminal
      ORIGINAL SOURCE:
        ORGANISM: Influenza virus
        INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
    FEATURE:
      NAME/KEY: AcNPV 61K protein signal sequence
      LOCATION: 1 to 18
    FEATURE:
      NAME/KEY: mature rHA
      LOCATION: 19 to 552
US-08-453-848-7
```

Query Match 2.9%; Score 7; DB 2; Length 570;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 378 SEGTGQA 384

```
RESULT 99
US-09-169-027-7
Sequence 7, Application US/09169027
Patent No. 6245532
GENERAL INFORMATION:
  APPLICANT: Smith, Gale Eugene
  APPLICANT: Volvovitz, Franklin
  APPLICANT: Wilkinson, Bethanie Eident
  APPLICANT: Voznesensky, Andrei I.
  APPLICANT: Hackett, Craig Stanway
  TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
  TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patrea L. Pabst
    STREET: 2800 One Atlantic Center
    STREET: 1201 West Peachtree Street
    CITY: Atlanta
    STATE: GA
    COUNTRY: USA
```

```
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/169,027
  FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/08/453,848
  FILING DATE: 30-MAY-1995
  APPLICATION NUMBER: 08/120,607
  FILING DATE: 13-SEPT-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Pabst, Patrea L.
  REGISTRATION NUMBER: 31,284
  REFERENCE/DOCKET NUMBER: MGS101CIP
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (404)-873-8794
    TELEFAX: (404)-873-8795
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 570 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE: N-terminal
      ORIGINAL SOURCE:
        ORGANISM: Influenza virus
        INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
    FEATURE:
      NAME/KEY: AcNPV 61K protein signal sequence
      LOCATION: 1 to 18
    FEATURE:
      NAME/KEY: mature rHA
      LOCATION: 19 to 552
US-09-169-027-7
```

Query Match 2.9%; Score 7; DB 3; Length 570;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 378 SEGTGQA 384

```
RESULT 100
US-08-453-848-15
Sequence 15, Application US/08453848
Patent No. 5858368
GENERAL INFORMATION:
  APPLICANT: Smith, Gale Eugene
  APPLICANT: Volvovitz, Franklin
  APPLICANT: Wilkinson, Bethanie Eident
  APPLICANT: Voznesensky, Andrei I.
  APPLICANT: Hackett, Craig Stanway
  TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
  TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patrea L. Pabst
    STREET: 2800 One Atlantic Center
    STREET: 1201 West Peachtree Street
    CITY: Atlanta
    STATE: GA
    COUNTRY: USA
    ZIP: 30309-3450
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,848
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/120,607
; FILING DATE: 13-SEPT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MGS101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORGANISM: Influenza virus
; INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
; FEATURE:
; NAME/KEY: AcNPV 61K protein signal sequence
; LOCATION: 1 to 18
; NAME/KEY: mature rHA
; LOCATION: 19 to 553
; US-08-453-848-15

```

```

Query Match      2.9%; Score 7; DB 2; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      165 SEGTGQA 171
      |||||
Db      379 SEGTGQA 385

```

```

RESULT 101
US-08-453-848-21
; Sequence 21, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,848
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/120,607
; FILING DATE: 13-SEPT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MGS101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORGANISM: Influenza virus
; INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA
; FEATURE:
; NAME/KEY: AcNPV 61K protein signal sequence
; LOCATION: 1 to 18
; NAME/KEY: mature rHA
; LOCATION: 19 to 569
; US-08-453-848-21

```

```

Query Match      2.9%; Score 7; DB 2; Length 571;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      165 SEGTGQA 171
      |||||
Db      379 SEGTGQA 385

```

```

RESULT 102
US-09-169-027-15
; Sequence 15, Application US/09169027
; Patent No. 6245532
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/169,027  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/453,848  
;; FILING DATE: 30-MAY-1995  
;; APPLICATION NUMBER: 08/120,607  
;; FILING DATE: 13-SEPT-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pabst, Patrea L.  
;; REGISTRATION NUMBER: 31,284  
;; REFERENCE/DOCKET NUMBER: MGS101CIP  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (404)-873-8794  
;; TELEFAX: (404)-873-8795  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 571 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE: N-terminal  
;; ORIGINAL SOURCE:  
;; ORGANISM: Influenza virus  
;; INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA  
;; FEATURE:  
;; NAME/KEY: AcNPV 61K protein signal sequence  
;; LOCATION: 1 to 18  
;; FEATURE:  
;; NAME/KEY: mature rHA  
;; LOCATION: 19 to 553  
;;  
US-09-169-027-15

Query Match 2.9%; Score 7; DB 3; Length 571;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 379 SEGTGQA 385

RESULT 103  
US-09-169-027-21  
; Sequence 21, Application US/09169027  
; Patent No. 6245532  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volnovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/169,027  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/453,848  
;; FILING DATE: 30-MAY-1995  
;; APPLICATION NUMBER: 08/120,607  
;; FILING DATE: 13-SEPT-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pabst, Patrea L.  
;; REGISTRATION NUMBER: 31,284  
;; REFERENCE/DOCKET NUMBER: MGS101CIP  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (404)-873-8794  
;; TELEFAX: (404)-873-8795  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 571 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE: N-terminal  
;; ORIGINAL SOURCE:  
;; ORGANISM: Influenza virus  
;; INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA  
;; FEATURE:  
;; NAME/KEY: AcNPV 61K protein signal sequence  
;; LOCATION: 1 to 18  
;; FEATURE:  
;; NAME/KEY: mature rHA  
;; LOCATION: 19 to 569  
;;  
US-09-169-027-21

Query Match 2.9%; Score 7; DB 3; Length 571;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 SEGTGQA 171  
Db 379 SEGTGQA 385

RESULT 104  
US-08-463-163-3  
; Sequence 3, Application US/08463163  
; Patent No. 5696237  
; GENERAL INFORMATION:  
; APPLICANT: Fitzgerald, David J.  
; APPLICANT: Chaudhary, Vijay K.  
; APPLICANT: Pastan, Ira H.  
; APPLICANT: Waldmann, Thomas A.  
; APPLICANT: Queen, Cary L.  
; TITLE OF INVENTION: Recombinant Antibody-Toxin Fusion Protein  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew  
; STREET: One Market Plaza, Steuart Street Tower  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,163  
; FILING DATE: 05-JUN-1995



```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/227,227
; FILING DATE: 22-JAN-1981
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/911,227
; FILING DATE: 24-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,361
; FILING DATE: 21-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,722
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-12211
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-163-3

Query Match 2.9%; Score 7; DB 1; Length 599;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 301 ALASPGS 307

RESULT 105
US-08-405-615-1
; Sequence 1, Application US/08405615
; Patent No. 5602095
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Fitzgerald, David J.
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
; TITLE OF INVENTION: Increased Activity
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ellen L. Weber
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,615
; FILING DATE:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,709
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 613 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-615-1

Query Match 2.9%; Score 7; DB 1; Length 613;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
Db 315 ALASPGS 321

RESULT 106
US-08-461-234-1
; Sequence 1, Application US/08461234
; Patent No. 5821238
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira H.
; APPLICANT: Fitzgerald, David J.
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
; TITLE OF INVENTION: Increased Activity
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,234
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,709
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,615
; FILING DATE: 15-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-36-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 613 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; APPLICATION TYPE: NO
; HYPOTHETICAL: NO
; US-08-461-234-1

Query Match 2.9%; Score 7; DB 2; Length 613;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177
```

Db 315 ALASPGS 321

RESULT 107

US-08-463-480-1

Sequence 1, Application US/08463480

Patent No. 5854044

GENERAL INFORMATION:

APPLICANT: Pastan, Ira H.

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,480

FILING DATE: 05-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/901,709

FILING DATE: 18-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,615

FILING DATE: 15-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen Lauver

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 15280-36-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 613 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-463-480-1

Query Match 2.9%; Score 7; DB 2; Length 613;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 171 ALASPGS 177

Db 315 ALASPGS 321

RESULT 108

US-09-479-479-2

Sequence 2, Application US/09479479

Patent No. 6423513

GENERAL INFORMATION:

APPLICANT:

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,479

FILING DATE: 06-NOV-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/297,851

FILING DATE: 30-JUL-1999

APPLICATION NUMBER: US 60/030,376

FILING DATE: 06-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/20207

FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 015280-29810US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 613 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-479-479-2

Query Match 2.9%; Score 7; DB 4; Length 613;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 171 ALASPGS 177

Db 315 ALASPGS 321

RESULT 109

US-09-297-851-2

Sequence 2, Application US/09297851

Patent No. 6426075

GENERAL INFORMATION:

APPLICANT:

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,479

FILING DATE: 06-NOV-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/297,851

FILING DATE: 30-JUL-1999

APPLICATION NUMBER: US 60/030,376

FILING DATE: 06-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/20207

FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 015280-29810US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 613 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-479-479-2

Query Match 2.9%; Score 7; DB 4; Length 613;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 171 ALASPGS 177

Db 315 ALASPGS 321

RESULT 109

US-09-297-851-2

Sequence 2, Application US/09297851

Patent No. 6426075

GENERAL INFORMATION:

APPLICANT:

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,479

FILING DATE: 06-NOV-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/297,851

FILING DATE: 30-JUL-1999

APPLICATION NUMBER: US 60/030,376

FILING DATE: 06-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/20207

FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 015280-29810US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 613 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-479-479-2

Query Match 2.9%; Score 7; DB 4; Length 613;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 171 ALASPGS 177

Db 315 ALASPGS 321

RESULT 109

US-09-297-851-2

Sequence 2, Application US/09297851

Patent No. 6426075

GENERAL INFORMATION:

APPLICANT:

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,479

FILING DATE: 06-NOV-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/297,851

FILING DATE: 30-JUL-1999

APPLICATION NUMBER: US 60/030,376

FILING DATE: 06-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/20207

FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 015280-29810US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 613 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-479-479-2

Query Match 2.9%; Score 7; DB 4; Length 613;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 171 ALASPGS 177

Db 315 ALASPGS 321

RESULT 109

US-09-297-851-2

Sequence 2, Application US/09297851

Patent No. 6426075

GENERAL INFORMATION:

APPLICANT:

ADDRESSEE: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,479

FILING DATE: 06-NOV-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/297,851

FILING DATE: 30-JUL-1999

APPLICATION NUMBER: US 60/030,376

FILING DATE: 06-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/20207

FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 015280-29810US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID

APPLICATION NUMBER: US/09/297,851  
FILING DATE: 30-JUL-1999  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/030,376  
FILING DATE: 06-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/20207  
FILING DATE: 05-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Storella, John R.  
REGISTRATION NUMBER: 32,944  
REFERENCE/DOCKET NUMBER: 015280-29810US  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 613 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-297-851-2

Query Match 2.9%; Score 7; DB 4; Length 613;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 315 ALASPGS 321

RESULT 110  
US-08-225-224-1  
Sequence 1, Application US/08225224  
Patent No. 5635599  
GENERAL INFORMATION:  
APPLICANT: PASTAN, Ira  
APPLICANT: KREITMAN, Robert J.  
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND  
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,224  
FILING DATE: 8-APR-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 15280-193  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 614 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..614  
OTHER INFORMATION: /label= native-PE  
US-08-225-224-1

Query Match 2.9%; Score 7; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 316 ALASPGS 322

RESULT 111  
US-08-722-258-1  
Sequence 1, Application US/08722258  
Patent No. 6011002  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Kreitman, Robert J.  
APPLICANT: Puri, Raj K.  
TITLE OF INVENTION: Circularly Permuted Ligands and  
TITLE OF INVENTION: Circularly Permuted Chimeric Molecules  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,258  
FILING DATE: 08-JAN-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/04468  
FILING DATE: 06-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/225,224  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 015280-193100US  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 614 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..614  
OTHER INFORMATION: /note= "native Pseudomonas exotoxin  
OTHER INFORMATION:  
US-08-722-258-1

Query Match 2.9%; Score 7; DB 3; Length 614;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 316 ALASPGS 322

RESULT 112  
PCT-US95-04468-1  
; Sequence 1, Application PC/TUS9504468  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND  
; TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS  
; NUMBER OF SEQUENCES: 59  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04468  
; FILING DATE: 07-APR-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/225,224  
; FILING DATE: 08-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen L.  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 15280-193-1PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 614 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..614  
; OTHER INFORMATION: /label= native-PE  
PCT-US95-04468-1

Query Match 2.9%; Score 7; DB 5; Length 614;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 316 ALASPGS 322

RESULT 113  
US-09-970-516-4  
; Sequence 4, Application US/09970516  
; Patent No. 6610534  
; GENERAL INFORMATION:  
; APPLICANT: No. 6610534artis AG  
; TITLE OF INVENTION: Induction of blood vessel formation through administration of  
; TITLE OF INVENTION: polynucleotides encoding sphingosine kinases  
; FILE REFERENCE: 4-31617  
; CURRENT APPLICATION NUMBER: US/09/970,516  
; CURRENT FILING DATE: 2001-10-04  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 4  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-516-4

Query Match 2.9%; Score 7; DB 4; Length 618;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 395 ALASPGS 401

RESULT 114  
US-08-356-786-16  
; Sequence 16, Application US/08356786  
; Patent No. 5877305  
; GENERAL INFORMATION:  
; APPLICANT: Huston, James S.  
; APPLICANT: Oppermann, Hermann  
; APPLICANT: Houston, L. L.  
; APPLICANT: Ring, David B.  
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer  
; TITLE OF INVENTION: Marker  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault  
; STREET: Exchange Place, 53 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/356,786  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/831,967  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: CRP-053  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 622 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-356-786-16

Query Match 2.9%; Score 7; DB 2; Length 622;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 ALASPGS 177  
Db 324 ALASPGS 330

RESULT 115  
US-09-046-992-2  
; Sequence 2, Application US/09046992  
; Patent No. 6140066  
; GENERAL INFORMATION:  
; APPLICANT: Lorberboun-Galski, Haya  
; APPLICANT: Yarkoni, Shai  
; APPLICANT: Ben-Yehudah, Ahmi  
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS

;; TITLE OF INVENTION: USING A CHIMERIC TOXIN  
;; NUMBER OF SEQUENCES: 7  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds, LLP  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036-2811  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: Windows  
;; SOFTWARE: FastSeq for Windows Version 2.0b  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/046,992  
;; FILING DATE: 24-MAR-1998  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Poissant, Brian M  
;; REGISTRATION NUMBER: 28,462  
;; REFERENCE/DOCKET NUMBER: 9457-0013-999  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650-493-4935  
;; TELEFAX: 650-493-5556  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 635 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: internal  
;; US-09-046-992-2

Query Match 2.9%; Score 7; DB 3; Length 635;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
|||  
Db 337 ALASPGS 343

RESULT 116  
US-08-235-838-14  
; Sequence 14, Application US/08235838  
; Patent No. 5571894  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/235,838  
; FILING DATE: TBA  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott

;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/235,838  
;; FILING DATE: TBA  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/828,832  
;; FILING DATE: 31-JAN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 91-810079.3  
;; FILING DATE: 05-FEB-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Elmer, James Scott  
;; REGISTRATION NUMBER: 36,129  
;; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (919)541-8614  
;; TELEFAX: (919)541-8689  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 637 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-235-838-14

Query Match 2.9%; Score 7; DB 1; Length 637;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
|||  
Db 339 ALASPGS 345

RESULT 117  
US-08-235-838-16  
; Sequence 16, Application US/08235838  
; Patent No. 5571894  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/235,838  
; FILING DATE: TBA  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-235-838-16

Query Match 2.9%; Score 7; DB 1; Length 637;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 339 ALASPGS 345

RESULT 118  
US-08-465-473B-14  
; Sequence 14, Application US/08465473B  
; Patent No. 5939531  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5939531man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NOVARTIS Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901-6940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/465,473B  
; FILING DATE: 5 June 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pfeiffer, Henna J.  
; REGISTRATION NUMBER: 22,640  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)522 6940  
; TELEFAX: (908)522 6955  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 637 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-473B-14

Query Match 2.9%; Score 7; DB 1; Length 637;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 339 ALASPGS 345

RESULT 120  
US-09-047-148-2  
; Sequence 2, Application US/09047148  
; Patent No. 6086900

Query Match 2.9%; Score 7; DB 2; Length 637;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 339 ALASPGS 345

RESULT 119  
US-08-465-473B-16  
; Sequence 16, Application US/08465473B  
; Patent No. 5939531  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5939531man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NOVARTIS Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901-6940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/465,473B  
; FILING DATE: 5 June 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pfeiffer, Henna J.  
; REGISTRATION NUMBER: 22,640  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)522 6940  
; TELEFAX: (908)522 6955  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 637 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-473B-16

Query Match 2.9%; Score 7; DB 2; Length 637;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 339 ALASPGS 345

RESULT 120  
US-09-047-148-2  
; Sequence 2, Application US/09047148  
; Patent No. 6086900

Matches	7;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	174	SPGSCLE	180						
Db	429	SPGSCLE	435						
RESULT 122									
US-08-398-590A-40									
; Sequence 40, Application US/08398590A									
; Patent No. 5935811									
; GENERAL INFORMATION:									
; APPLICANT: Anderson, David J.									
; APPLICANT: Schoenherr, Christopher J.									
; TITLE OF INVENTION: Neuron-Restrictive Silencer Factor									
; TITLE OF INVENTION: Proteins									
; NUMBER OF SEQUENCES: 53									
; CORRESPONDENCE ADDRESS:									
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert									
; STREET: Four Embarcadero Center, Suite 3400									
; CITY: San Francisco									
; STATE: California									
; COUNTRY: United States									
; ZIP: 94111-4187									
; COMPUTER READABLE FORM:									
; MEDIUM TYPE: Floppy disk									
; COMPUTER: IBM PC compatible									
; OPERATING SYSTEM: PC-DOS/MS-DOS									
; SOFTWARE: PatentIn Release #1.0, Version #1.30									
; CURRENT APPLICATION DATA:									
; APPLICATION NUMBER: US/08/398,590A									
; FILING DATE: 03-MAR-1995									
; CLASSIFICATION: 435									
; PRIOR APPLICATION DATA:									
; APPLICATION NUMBER: US 08/103,445									
; FILING DATE: 06-AUG-1995									
; ATTORNEY/AGENT INFORMATION:									
; NAME: Silva, Robin M.									
; REGISTRATION NUMBER: 38,304									
; REFERENCE/DOCKET NUMBER: A-60897/RFT/RMS									
; TELECOMMUNICATION INFORMATION:									
; TELEPHONE: (415) 781-1989									
; TELEFAX: (415) 398-3249									
; TELEX: 910 277299									
; INFORMATION FOR SEQ ID NO: 40:									
; SEQUENCE CHARACTERISTICS:									
; LENGTH: 676 amino acids									
; TYPE: amino acid									
; TOPOLOGY: linear									
; MOLECULE TYPE: protein									
US-08-398-590A-40									
Query Match 2.9%; Score 7; DB 2; Length 676;									
Best Local Similarity 100.0%; Pred. No. 97;									
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	7	GDGSGPA	13						
Db	64	GDGSGPA	70						
RESULT 123									
US-08-894-997-40									
; Sequence 40, Application US/08894997A									
; Patent No. 6270990									
; GENERAL INFORMATION:									
; APPLICANT: Anderson, David J									
; APPLICANT: Schoenherr, Christopher J									
; TITLE OF INVENTION: NEURON-RESTRICTIVE SILENCER FACTOR									
; FILE REFERENCE: 17810-502 NRSF									
; CURRENT APPLICATION NUMBER: US/08/894,997A									
; CURRENT FILING DATE: 1998-01-06									
; EARLIER APPLICATION NUMBER: PCT/US96/02817									

; EARLIER FILING DATE: 1996-03-01  
; EARLIER APPLICATION NUMBER: 08/398,590  
; EARLIER FILING DATE: 1995-03-03  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 40  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CHAIN  
; LOCATION: (1)..(676)  
; OTHER INFORMATION: Human NSRF (partial)  
US-08-894-997-40

Query Match 2.9%; Score 7; DB 3; Length 676;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDSGSPA 13  
Db 64 GDSGSPA 70

## RESULT 124

US-09-252-991A-19361  
; Sequence 19361, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19361  
; LENGTH: 767  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19361

Query Match 2.9%; Score 7; DB 4; Length 767;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 ALASPGS 177  
Db 469 ALASPGS 475

## RESULT 125

US-08-494-168-2  
; Sequence 2, Application US/08494168  
; Patent No. 5731192  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T.  
; APPLICANT: Zhou, Jing  
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method  
; TITLE OF INVENTION: of Detecting Collagen Deficiency  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/494,168  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/112,465  
; FILING DATE: 27-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 40397/104/BARR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1694 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-494-168-2

Query Match 2.9%; Score 7; DB 1; Length 1694;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 SRCQVCV 240  
Db 1676 SRCQVCV 1682

## RESULT 126

US-09-679-279-14  
; Sequence 14, Application US/09679279  
; Patent No. 6524841  
; GENERAL INFORMATION:  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Volchegursky, Yanina  
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
; TITLE OF INVENTION: Genes and Uses Thereof  
; FILE REFERENCE: 300622004700  
; CURRENT APPLICATION NUMBER: US/09/679,279  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/158,305  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: US 60/190,024  
; PRIOR FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 3562  
; TYPE: PRT  
; ORGANISM: Micromonospora megalomicea  
US-09-679-279-14

Query Match 2.9%; Score 7; DB 4; Length 3562;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 GTVPLYS 43  
Db 2228 GTVPLYS 2234

## RESULT 127

US-09-439-897-51  
; Sequence 51, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:



APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 95-1263-C  
CURRENT APPLICATION NUMBER: US/09/439,897  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 51  
LENGTH: 6  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
OTHER INFORMATION: construct C1 alpha3  
US-09-439-897-51

Query Match 2.5%; Score 6; DB 3; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATWTTT 13  
Db 1 ATWTTT 6

RESULT 128  
US-10-080-505-58  
Sequence 58, Application US/10080505  
Patent No. 6676948  
GENERAL INFORMATION:  
APPLICANT: St. Geme, Joseph W.  
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
FILE REFERENCE: A-59941-1/RFT/DCF/DHR  
CURRENT APPLICATION NUMBER: US/10/080,505  
CURRENT FILING DATE: 2002-02-22  
PRIOR APPLICATION NUMBER: US 08/296,791  
PRIOR FILING DATE: 1994-10-25  
PRIOR APPLICATION NUMBER: US 09/839,996  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 58  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-10-080-505-58

Query Match 2.5%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12  
Db 1 GDSGSP 6

RESULT 129  
US-08-296-791-7  
Sequence 7, Application US/08296791  
Patent No. 6245337  
GENERAL INFORMATION:  
APPLICANT: St. Geme III, Joseph W.  
APPLICANT: Falkow, Stanley  
TITLE OF INVENTION: Haemophilus Adherence and Penetration  
TITLE OF INVENTION: Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tracartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-296-791-7

Query Match 2.5%; Score 6; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GDSGSP 12  
Db 1 GDSGSP 6

RESULT 130  
US-08-296-791-8  
Sequence 8, Application US/08296791  
Patent No. 6245337  
GENERAL INFORMATION:  
APPLICANT: St. Geme III, Joseph W.  
APPLICANT: Falkow, Stanley  
TITLE OF INVENTION: Haemophilus Adherence and Penetration  
TITLE OF INVENTION: Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tracartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-296-791-8

```
Query Match      2.5%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GDSGSP 12
      |||||
Db      1 GDSGSP 6

RESULT 131
US-09-839-996-7
; Sequence 7, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; PROTEIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-839-996-7

Query Match      2.5%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GDSGSP 12
      |||||
Db      1 GDSGSP 6

RESULT 132
US-09-839-996-8
; Sequence 8, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; PROTEIN
```

```
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-839-996-8

Query Match      2.5%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GDSGSP 12
      |||||
Db      1 GDSGSP 6

RESULT 133
US-10-080-505-53
; Sequence 53, Application US/10080505
; Patent No. 6678948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-53

Query Match      2.5%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GDSGSP 12
      |||||
```

Db 1 GDSGSP 6

## RESULT 134

US-10-080-505-54  
; Sequence 54, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFT/DC7/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 54  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-080-505-54

Query Match 2.5%; Score 6; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;  
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

Qy 7 GDSGSP 12

Db 1 GDSGSP 6

## RESULT 135

PCT-US95-10661A-7  
; Sequence 7, Application PC/TUS9510661A  
; GENERAL INFORMATION:  
; APPLICANT: Washington University, et al.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10661A  
; FILING DATE: 16-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: FP-59941/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 398-3249  
; TELEFAX: (415) 781-1989  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
PCT-US95-10661A-7

Query Match 2.5%; Score 6; DB 5; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;  
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

Qy 7 GDSGSP 12

Db 1 GDSGSP 6

## RESULT 136

PCT-US95-10661A-8  
; Sequence 8, Application PC/TUS9510661A  
; GENERAL INFORMATION:  
; APPLICANT: Washington University, et al.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10661A  
; FILING DATE: 16-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: FP-59941/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
PCT-US95-10661A-8

Query Match 2.5%; Score 6; DB 5; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;  
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

Qy 7 GDSGSP 12

Db 1 GDSGSP 6

## RESULT 137

US-09-322-624-4  
; Sequence 4, Application US/09322624  
; Patent No. 6548734  
; GENERAL INFORMATION:  
; APPLICANT: Glimcher, L et al.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO MODULATION OF  
; CARYLLAGE GROWTH BY MODULATION OF NFATP ACTIVITY  
; FILE REFERENCE: HUI-035CP  
; CURRENT APPLICATION NUMBER: US/09/322,624

; CURRENT FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: USSN 09/087,139  
; EARLIER FILING DATE: 1998-03-28  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-322-624-4

Query Match 2.5%; Score 6; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 PAIAIA 133  
Db 2 PAIAIA 7

RESULT 138  
US-10-080-505-20  
; Sequence 20, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-080-505-20

Query Match 2.5%; Score 6; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GDSGSP 12  
Db 1 GDSGSP 6

RESULT 139  
US-08-260-582-40  
; Sequence 40, Application US/08260582  
; Patent No. 5635182  
; GENERAL INFORMATION:  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lu, Zhijian  
; TITLE OF INVENTION: METHOD OF DETECTING LIGAND  
; TITLE OF INVENTION: INTERACTIONS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/260,582  
; FILING DATE: 16-JUN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meinerdt, M. C.  
; REGISTRATION NUMBER: 31,544  
; REFERENCE/DOCKET NUMBER: GI 5236  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 876-1170  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-260-582-40

Query Match 2.5%; Score 6; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TQQAAL 173  
Db 3 TQQAAL 8

RESULT 140  
PCT-US95-05471-40  
; Sequence 40, Application PC/TUS9505471  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHOD OF DETECTING LIGAND INTERACTIONS  
; NUMBER OF SEQUENCES: 76  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/05471  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US95-05471-40

Query Match 2.5%; Score 6; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TQQAAL 173  
Db 3 TQQAAL 8

RESULT 141  
US-09-439-897-58  
; Sequence 58, Application US/09439897  
; Patent No. 6277558  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Billy G  
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
; FILE REFERENCE: 95-1263-C

```

; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 58
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: construct C5 alpha
US-09-439-897-58

Query Match          2.5%; Score 6; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MAPITG 112
Db 6 MAPITG 11

RESULT 142
US-09-106-568E-149
; Sequence 149, Application US/09106568E
; Patent No. 6455248
; GENERAL INFORMATION:
; APPLICANT: Bhattacherjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacherjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; TITLE OF INVENTION: A BIOLOGICAL SAMPLE
; FILE REFERENCE: 96,247-A
; CURRENT APPLICATION NUMBER: US/09/106,568E
; CURRENT FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 149
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide segment of LYS2_CaLB shown in Figure 4.
US-09-106-568E-149

Query Match          2.5%; Score 6; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 TTAIPS 33
Db 10 TTAIPS 15

RESULT 143
US-08-716-249-1
; Sequence 1, Application US/08716249
; Patent No. 6455244
; GENERAL INFORMATION:
; APPLICANT: Guichard, Gilles
; APPLICANT: Muller, Sylviane
; APPLICANT: Briand, Jean-Paul
; APPLICANT: Regemortel, Marc
; TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
; TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spencer & Frank
; STREET: 1100 New York Avenue, Suite 300E
; CITY: Washington, D.C.
; COUNTRY: USA

```

```

; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,249
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/00292
; FILING DATE: 13-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Calvetti, Frederick F.
; REGISTRATION NUMBER: 28,557
; REFERENCE/DOCKET NUMBER: GROFO 7001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)414-4000
; TELEFAX: (202)414-4040
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-716-249-1

Query Match          2.5%; Score 6; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSG 11
Db 6 RGDGSG 11

RESULT 144
US-08-716-249-9
; Sequence 9, Application US/08716249
; Patent No. 6455244
; GENERAL INFORMATION:
; APPLICANT: Guichard, Gilles
; APPLICANT: Muller, Sylviane
; APPLICANT: Briand, Jean-Paul
; APPLICANT: Regemortel, Marc
; TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
; TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spencer & Frank
; STREET: 1100 New York Avenue, Suite 300E
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,249
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/00292
; FILING DATE: 13-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Calvetti, Frederick F.
; REGISTRATION NUMBER: 28,557
; REFERENCE/DOCKET NUMBER: GROFO 7001

```

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)414-4000  
TELEFAX: (202)414-4040  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES  
ANTI-SENSE: YES  
US-08-716-249-9

Query Match 2.5%; Score 6; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RGDGSGS 11

Db 6 RGDGSGS 11

## RESULT 145

US-08-449-287-20  
Sequence 20, Application US/08449287  
Patent No. 5877293

## GENERAL INFORMATION:

APPLICANT: ADAIR, John Robert  
APPLICANT: BODMER, Mark William  
APPLICANT: MOUNTAIN, Andrew  
APPLICANT: OWENS, Raymond John  
TITLE OF INVENTION: CDR Grafted Anti-CEA Antibodies and  
TITLE OF INVENTION: Their Production  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,287  
FILING DATE:  
CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/154,389

## FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT GB91/01108

FILING DATE: 05-JUL-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9014932.9

FILING DATE: 05-JUL-1990

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT GB90/02017

FILING DATE: 21-DEC-1990

## ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D.

REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 40283/110 CARA

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

## INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

## TELECOMMUNICATION INFORMATION:

US-08-449-287-20

## Query Match

2.5%; Score 6; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134

Db 5 AIAIAV 10

## RESULT 146

US-09-003-081-8  
Sequence 8, Application US/09003081  
Patent No. 5988779

## GENERAL INFORMATION:

APPLICANT: Campfield, Arthur Dr.  
APPLICANT: Devos, Rene Dr.  
APPLICANT: Guisez, Yves Dr.  
TITLE OF INVENTION: Recombinant Obese (OB) Proteins  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche, Inc.  
STREET: 340 Kingaland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/003,081  
FILING DATE:  
CLASSIFICATION: 514

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/435,777

## FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Picut, Catherine A

REGISTRATION NUMBER: 37419

REFERENCE/DOCKET NUMBER: 9165

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-4387

TELEFAX: (201) 235-2363

## INFORMATION FOR SEQ ID NO: 8:

## SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

## TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

## US-09-003-081-8

## Query Match

2.5%; Score 6; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134

Db 5 AIAIAV 10

## RESULT 147

US-08-648-262-8  
Sequence 8, Application US/08648262  
Patent No. 6025324

## GENERAL INFORMATION:

APPLICANT: Bailon, Pascal Mr.  
APPLICANT: Campfield, Arthur Dr.  
APPLICANT: Devos, Rene Dr.  
APPLICANT: Guisez, Yves Dr.  
TITLE OF INVENTION: Pegylated Obese (OB) Proteins  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche, Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/648.262  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Epstein, William H.  
REGISTRATION NUMBER: 20008  
REFERENCE/DOCKET NUMBER: 9281  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-3723  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-648-262-8

Query Match 2.5%; Score 6; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134  
DB 5 AIAIAV 10

RESULT 148  
US-08-648-263-8  
Sequence 8, Application US/08/648.263  
Patent No. 6025325  
GENERAL INFORMATION:  
APPLICANT: Campfield, Arthur  
APPLICANT: Devos, Rene  
APPLICANT: Guisez, Yves  
TITLE OF INVENTION: RECOMBINANT OBSE (OB) PROTEINS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/648.263

FILING DATE: 15-MAY-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/484,629  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/435,777  
FILING DATE: 05-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kreisler, Lewis J.  
REGISTRATION NUMBER: 38522  
REFERENCE/DOCKET NUMBER: RAN 4105/175-002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-4387  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-648-263-8

Query Match 2.5%; Score 6; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134  
DB 5 AIAIAV 10

RESULT 149  
US-08-840-713-48  
Sequence 48, Application US/08840713  
Patent No. 6498233  
GENERAL INFORMATION:  
APPLICANT: WELS, Winfried, Dr.  
APPLICANT: FOYMINAVA, Jesus  
TITLE OF INVENTION: NUCLEIC ACID TRANSFER SYSTEM  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaide, Marmelstein, Murray & Oram LLP  
STREET: 655 15th St., N.W., Suite 330 - G St. Lobby  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/840.713  
FILING DATE: 25-APR-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitts, Monica Chin  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: 1614-7014  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638 - 5000  
TELEFAX: (202) 638 - 4810  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-840-713-48

Query Match 2.5%; Score 6; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134  
Db 5 AIAIAV 10

RESULT 150

US-09-904-196B-3  
; Sequence 3, Application US/09904196B  
; Patent No. 6555660  
; GENERAL INFORMATION:  
; APPLICANT: NISSEN, TORBEN LAURSGAARD  
; APPLICANT: ANDERSEN, KIM VILBOUR  
; APPLICANT: HANSEN, CHRISTIAN KARSTEN  
; APPLICANT: MIKKESEN, JAN MOLLER  
; TITLE OF INVENTION: G-CSF CONJUGATES  
; FILE REFERENCE: 31-000700US  
; CURRENT FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: US/09/904,196B  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US/09/760,008  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: 60/176,376  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: 60/189,506  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: 60/215,644  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DK PA 2000 00024  
; PRIOR FILING DATE: 2000-01-10  
; PRIOR APPLICATION NUMBER: DK PA 2000 00341  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: DK PA 2000 00943  
; PRIOR FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-904-196B-3

Query Match 2.5%; Score 6; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 AIAIAV 134  
Db 5 AIAIAV 10

Search completed: April 5, 2004, 07:39:59  
Job time : 26 secs